



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 125446**

**TO: Manjunath N Rao**  
**Location: rem/3b81/3c70**  
**Art Unit: 1652**  
**Monday, June 28, 2004**

**Case Serial Number: 10/038723**

**From: Alex Waclawiw**  
**Location: Biotech-Chem Library**  
**Rem 1A71**  
**Phone: 272-2534**

**Alexandra.waclawiw@uspto.gov**

### **Search Notes**

**THIS PAGE LEFT BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 07:44:58 ; Search time 50 Seconds  
(without alignments)  
3015.110 Million cell updates/sec

Title: US-10-038-723-2  
Perfect score: 2771  
Sequence: 1 MSFRLALSLGLVCTGLANV.....SKTTATASKSTTTTBSGMSL 534

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : Published Applications AA.\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2771	100.0	534	9	US-09-908-395-2
2	2771	100.0	534	10	US-09-821-616-9
3	2771	100.0	534	14	US-10-038-723-2
4	2771	100.0	534	15	US-10-421-586-2
5	2771	100.0	534	15	US-10-421-586-3
6	2742	99.0	640	14	US-10-038-723-13
7	2742	99.0	640	15	US-10-421-586-14
8	2735	98.7	743	15	US-10-418-836-9
9	2735	98.7	743	15	US-10-418-836-13
10	2735	98.7	743	15	US-10-418-836-10
11	2735	98.7	743	15	US-10-418-836-16
12	2720.5	98.2	748	15	US-10-418-836-19
13	2716	98.0	741	15	US-10-418-836-30
14	2715.5	98.0	972	15	US-10-418-836-38
15	2715	98.0	739	15	US-10-418-836-26

16	2715	98.0	739	15	US-10-418-836-27	Sequence 27, Appl
17	2714.5	98.0	740	15	US-10-418-836-28	Sequence 28, Appl
18	2714.5	98.0	740	15	US-10-418-836-34	Sequence 34, Appl
19	2714.5	98.0	742	15	US-10-418-836-29	Sequence 29, Appl
20	2714	97.9	743	15	US-10-418-836-31	Sequence 31, Appl
21	2714	97.9	743	15	US-10-418-836-35	Sequence 35, Appl
22	2714	97.9	743	15	US-10-418-836-39	Sequence 39, Appl
23	2523	91.1	1095	15	US-10-228-063-45	Sequence 45, Appl
24	1850.5	66.8	631	14	US-10-213-990-54	Sequence 34, Appl
25	1689	61.0	618	10	US-09-821-616-34	Sequence 7, Appl
26	1674.5	60.4	591	10	US-09-821-616-7	Sequence 3109, Ap
27	1548.5	55.9	647	15	US-10-369-493-3109	Sequence 3, Appl
28	1425.5	51.4	581	14	US-10-281-673-3	Sequence 21, Appl
29	1408.5	50.8	704	14	US-10-213-990-21	Sequence 33, Appl
30	1380.5	49.8	620	14	US-10-213-990-33	Sequence 12402, A
31	1250	45.1	441	15	US-10-369-493-12402	Sequence 12533, A
32	1132	40.9	432	15	US-10-369-493-12533	Sequence 2490, Ap
33	781.5	28.2	450	15	US-10-369-493-2490	Sequence 49, Appl
34	725	26.2	579	14	US-10-228-063-49	Sequence 22096, A
35	597	21.5	549	15	US-10-369-493-22096	Sequence 12713, A
36	164.5	5.9	2344	9	US-09-815-242-12713	Sequence 4, Appl
37	160	5.8	35	10	US-09-821-616-4	Sequence 43924, A
38	159.5	5.8	2283	14	US-10-172-502-4	Sequence 22616, A
39	159	5.7	2271	12	US-10-282-122A-43924	Sequence 11, Appl
40	158	5.7	1283	15	US-10-369-493-22616	Sequence 10932, A
41	155.5	5.6	1621	14	US-10-185-990-11	Sequence 63, Appl
42	155.5	5.6	1621	14	US-10-185-990-10	Sequence 53, Appl
43	154.5	5.6	1031	9	US-09-815-242-10932	Sequence 1555, Ap
44	152.5	5.5	45	15	US-10-120-801-63	Sequence 525, App
45	152	5.5	45	15	US-10-120-801-63	Sequence 525, App
46	151.5	5.5	1099	12	US-10-282-122A-48220	Sequence 525, App
47	150	5.4	1099	12	US-10-369-493-1535	Sequence 525, App
48	147.5	5.3	4440	12	US-10-206-915-525	Sequence 525, App
49	147.5	5.3	4440	12	US-10-199-670-525	Sequence 525, App
50	147.5	5.3	4440	12	US-10-201-858-525	Sequence 525, App
51	147.5	5.3	4440	12	US-10-205-890-525	Sequence 525, App
52	147.5	5.3	4440	12	US-10-208-024-525	Sequence 525, App
53	147.5	5.3	4440	12	US-10-201-853-525	Sequence 525, App
54	147.5	5.3	4440	12	US-10-171-581-525	Sequence 525, App
55	147.5	5.3	4440	12	US-10-176-483-525	Sequence 525, App
56	147.5	5.3	4440	12	US-10-176-483-525	Sequence 525, App
57	147.5	5.3	4440	12	US-10-176-749-525	Sequence 525, App
58	147.5	5.3	4440	12	US-10-176-914-525	Sequence 525, App
59	147.5	5.3	4440	12	US-10-176-915-525	Sequence 525, App
60	147.5	5.3	4440	12	US-10-180-550-525	Sequence 525, App
61	147.5	5.3	4440	12	US-10-183-014-525	Sequence 525, App
62	147.5	5.3	4440	12	US-10-187-728-525	Sequence 525, App
63	147.5	5.3	4440	12	US-10-187-740-525	Sequence 525, App
64	147.5	5.3	4440	12	US-10-187-883-525	Sequence 525, App
65	147.5	5.3	4440	12	US-10-194-353-525	Sequence 525, App
66	147.5	5.3	4440	12	US-10-194-460-525	Sequence 525, App
67	147.5	5.3	4440	12	US-10-194-463-525	Sequence 525, App
68	147.5	5.3	4440	12	US-10-194-484-525	Sequence 525, App
69	147.5	5.3	4440	12	US-10-195-884-525	Sequence 525, App
70	147.5	5.3	4440	12	US-10-195-896-525	Sequence 525, App
71	147.5	5.3	4440	12	US-10-196-744-525	Sequence 525, App
72	147.5	5.3	4440	12	US-10-196-755-525	Sequence 525, App
73	147.5	5.3	4440	12	US-10-196-755-525	Sequence 525, App
74	147.5	5.3	4440	12	US-10-197-704-525	Sequence 525, App
75	147.5	5.3	4440	12	US-10-197-710-525	Sequence 525, App
76	147.5	5.3	4440	12	US-10-197-710-525	Sequence 525, App
77	147.5	5.3	4440	12	US-10-198-758-525	Sequence 525, App
78	147.5	5.3	4440	12	US-10-198-766-525	Sequence 525, App
79	147.5	5.3	4440	12	US-10-199-304-525	Sequence 525, App
80	147.5	5.3	4440	12	US-10-199-309-525	Sequence 525, App
81	147.5	5.3	4440	12	US-10-199-313-525	Sequence 525, App
82	147.5	5.3	4440	12	US-10-199-456-525	Sequence 525, App
83	147.5	5.3	4440	12	US-10-201-319-525	Sequence 525, App
84	147.5	5.3	4440	12	US-10-202-412-525	Sequence 525, App
85	147.5	5.3	4440	12	US-10-206-919-525	Sequence 525, App
86	147.5	5.3	4440	12	US-10-206-922-525	Sequence 525, App
87	147.5	5.3	4440	12	US-10-206-924-525	Sequence 525, App
88	147.5	5.3	4440	12	US-10-206-928-525	Sequence 525, App
					US-10-207-914-525	Sequence 525, App

89 147.5 5.3 4440 12 US-10-207-921-525 Sequence 525, App  
90 147.5 5.3 4440 12 US-10-207-922-525 Sequence 525, App  
91 147.5 5.3 4440 12 US-10-208-027-525 Sequence 525, App  
92 147.5 5.3 4440 12 US-10-174-570-525 Sequence 525, App  
93 147.5 5.3 4440 12 US-10-183-005-525 Sequence 525, App  
94 147.5 5.3 4440 13 US-10-052-586-525 Sequence 525, App  
95 147.5 5.3 4440 14 US-10-174-590-525 Sequence 525, App  
96 147.5 5.3 4440 14 US-10-176-758-525 Sequence 525, App  
97 147.5 5.3 4440 14 US-10-175-737-525 Sequence 525, App  
98 147.5 5.3 4440 14 US-10-173-706-525 Sequence 525, App  
99 147.5 5.3 4440 14 US-10-175-738-525 Sequence 525, App  
100 147.5 5.3 4440 14 US-10-175-752-525 Sequence 525, App

## ALIGNMENTS

RESULT 1  
US-09-908-395-2  
Sequence 2, Application US/09908395  
Patent No. US20020164723A1  
GENERAL INFORMATION:  
APPLICANT: Liaw, Gin  
APPLICANT: Pedersen, Sven  
TITLE OF INVENTION: A Method of Producing Saccharide  
FILE REFERENCE: 5318.200-US  
CURRENT APPLICATION NUMBER: US/09/908,395  
CURRENT FILING DATE: 2001-07-18  
PRIOR APPLICATION NUMBER: US/09/198,672  
PRIOR FILING DATE: PRIOR FILING DATE: 1998-11-23  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 534  
TYPE: PRT  
ORGANISM: Aspergillus Niger  
FEATURE:  
NAME/KEY: SIGNAL  
LOCATION: (1)...(24)  
US-09-908-395-2

Query Match 100.0%; Score 2771; DB 9; Length 534;  
Best Local Similarity 100.0%; Pred. No. 2,7e-237;  
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSFRSLALSGLVCTGTGLANVISKRATLDSWLSNEATVARTAILNINIGADGAWVSGADSGI 60  
DB 1 MSFRSLALSGLVCTGTGLANVISKRATLDSWLSNEATVARTAILNINIGADGAWVSGADSGI 60  
QY 61 VVASPTDNDPYFTYTRDSGLVTKTVLDFRNGDTSLSLTENYISAQAIYVGISNPSG 120  
DB 61 VVASPTDNDPYFTYTRDSGLVTKTVLDFRNGDTSLSLTENYISAQAIYVGISNPSG 120  
QY 121 DLSGAGLGPKEVNDETAAYTGSWGRPQDGPALRATAMIGFGQMLDNGYSTATDIW 180  
DB 121 DLSGAGLGPKEVNDETAAYTGSWGRPQDGPALRATAMIGFGQMLDNGYSTATDIW 180  
QY 121 DLSGAGLGPKEVNDETAAYTGSWGRPQDGPALRATAMIGFGQMLDNGYSTATDIW 180  
DB 121 DLSGAGLGPKEVNDETAAYTGSWGRPQDGPALRATAMIGFGQMLDNGYSTATDIW 180  
QY 181 PLVRNDLSYVAQYWNQGYDLMEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDQ 240  
DB 181 PLVRNDLSYVAQYWNQGYDLMEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDQ 240  
QY 241 APEILCYLQSFMTGSLFLANFDSRSRGKANTLLGSIHTFDPEAACDSTFQPCSPRALA 300  
DB 241 APEILCYLQSFMTGSLFLANFDSRSRGKANTLLGSIHTFDPEAACDSTFQPCSPRALA 300  
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAGRYPEDTYNGNPMFLCTTLAAEQLYDALYQMD 360  
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAGRYPEDTYNGNPMFLCTTLAAEQLYDALYQMD 360  
QY 361 KQGSLEVTDVSLDFFKALYSDDATGYSSSSSTYSIVDAVKTFADGFSIVETHAASNG 420  
DB 361 KQGSLEVTDVSLDFFKALYSDDATGYSSSSSTYSIVDAVKTFADGFSIVETHAASNG 420

DB 361 KQGSLEVTDVSLDFFKALYSDDATGYSSSSSTYSIVDAVKTFADGFSIVETHAASNG 420  
QY 421 SMSEQYDKSDGEOBSARDLWTSYALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480  
DB 421 SMSEQYDKSDGEOBSARDLWTSYALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480  
QY 481 TVSSVTVTSWPSIVATGCTTTATPTGSGSVTSTKTTATASKTSTTTSGMSL 534  
DB 481 TVSSVTVTSWPSIVATGCTTTATPTGSGSVTSTKTTATASKTSTTTSGMSL 534

RESULT 2  
US-09-821-616-9  
Sequence 9, Application US/09821616  
Publication No. US20030027290A1  
GENERAL INFORMATION:  
APPLICANT: Nielsen, Bjarne R.  
APPLICANT: Nielsen, Ruby  
TITLE OF INVENTION: Thermostable Glucoamylase  
FILE REFERENCE: 5279.200-US  
CURRENT APPLICATION NUMBER: US/09/821,616  
CURRENT FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/199,290  
PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-24  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 1557/97  
PRIOR FILING DATE: EARLIER FILING DATE: 1997-12-30  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 0925/98  
PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-10  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/070,746  
PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-08  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/094,344  
PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-28  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/979,673  
PRIOR FILING DATE: EARLIER FILING DATE: 1997-11-26  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/107,657  
PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-30  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 9  
LENGTH: 534  
TYPE: PRT  
ORGANISM: Aspergillus niger  
FEATURE:  
NAME/KEY: SIGNAL  
LOCATION: (1)...(24)  
US-09-821-616-9

Query Match 100.0%; Score 2771; DB 10; Length 534;  
Best Local Similarity 100.0%; Pred. No. 2,7e-237;  
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSFRSLALSGLVCTGTGLANVISKRATLDSWLSNEATVARTAILNINIGADGAWVSGADSGI 60  
DB 1 MSFRSLALSGLVCTGTGLANVISKRATLDSWLSNEATVARTAILNINIGADGAWVSGADSGI 60  
QY 61 VVASPTDNDPYFTYTRDSGLVTKTVLDFRNGDTSLSLTENYISAQAIYVGISNPSG 120  
DB 61 VVASPTDNDPYFTYTRDSGLVTKTVLDFRNGDTSLSLTENYISAQAIYVGISNPSG 120  
QY 121 DLSGAGLGPKEVNDETAAYTGSWGRPQDGPALRATAMIGFGQMLDNGYSTATDIW 180  
DB 121 DLSGAGLGPKEVNDETAAYTGSWGRPQDGPALRATAMIGFGQMLDNGYSTATDIW 180  
QY 121 DLSGAGLGPKEVNDETAAYTGSWGRPQDGPALRATAMIGFGQMLDNGYSTATDIW 180  
DB 121 DLSGAGLGPKEVNDETAAYTGSWGRPQDGPALRATAMIGFGQMLDNGYSTATDIW 180  
QY 181 PLVRNDLSYVAQYWNQGYDLMEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDQ 240  
DB 181 PLVRNDLSYVAQYWNQGYDLMEVNGSSFFTIAVQHRALVEGSAFATAVGSSCSWCDQ 240  
QY 241 APEILCYLQSFMTGSLFLANFDSRSRGKANTLLGSIHTFDPEAACDSTFQPCSPRALA 300  
DB 241 APEILCYLQSFMTGSLFLANFDSRSRGKANTLLGSIHTFDPEAACDSTFQPCSPRALA 300  
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAGRYPEDTYNGNPMFLCTTLAAEQLYDALYQMD 360  
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAGRYPEDTYNGNPMFLCTTLAAEQLYDALYQMD 360



```
Db 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRPEPTYNGNPMFLCTIAAEQYDALYQWD 360
Qy 361 KQGSLEVTVDLDFPFAALYSDAATGYSSSSSTYSSIVDAVKTFADGPFVSVETHAASNG 420
Db 361 KQGSLEVTVDLDFPFAALYSDAATGYSSSSSTYSSIVDAVKTFADGPFVSVETHAASNG 420
Qy 421 SMSEQYDKSDGDEQLSARDLTWGSYAALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
Db 421 SMSEQYDKSDGDEQLSARDLTWGSYAALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
Qy 481 TYSSVTVTWSPSIIVATGCTTTTATPTGSGSVTSTKTTATASKTSTTTTSGMSL 534
Db 481 TYSSVTVTWSPSIIVATGCTTTTATPTGSGSVTSTKTTATASKTSTTTTSGMSL 534

RESULT 3
US-10-038-723-2
; Sequence 2, Application US/10038723
; Publication No. US20030032163A1
; GENERAL INFORMATION:
; APPLICANT: Nielsen, Bjarne Roenfeldt
; APPLICANT: Svendsen, Allan
; APPLICANT: Pedersen, Henrik
; APPLICANT: Vind, Jesper
; APPLICANT: Hendriksen, Hanne Vang
; APPLICANT: Frandsen, Torben Peter
; TITLE OF INVENTION: Glucoamylase Variants
; FILE REFERENCE: 5636.200-US
; CURRENT APPLICATION NUMBER: US/10/038,723
; CURRENT FILING DATE: 2002-01-02
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/351,814
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PA 1998 00937
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PA 1998 01667
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/093,528
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/115,545
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-10-038-723-2

Query Match 100.0%; Score 2771; DB 14; Length 534;
Best Local Similarity 100.0%; Pred. No. 2,7e-237;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSFRSLALSGLVCTGLANVISKRATLDLSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
Db 1 MSFRSLALSGLVCTGLANVISKRATLDLSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
Qy 61 VVASPSTNDPDEYTYTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIVQGISNPSG 120
Db 61 VVASPSTNDPDEYTYTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIVQGISNPSG 120
Qy 121 DLSSGAGLEPKKNVDETAYTGSWGRPDGPALRATAMIGFQWLLDNGYSTTADIW 180
Db 121 DLSSGAGLEPKKNVDETAYTGSWGRPDGPALRATAMIGFQWLLDNGYSTTADIW 180
Qy 181 PLVRNDLSYVAQYWNQGTGYDLMEEVNGSFFTTIAVQHRALVEGSAFATAVSSCSWCDSQ 240
Db 181 PLVRNDLSYVAQYWNQGTGYDLMEEVNGSFFTTIAVQHRALVEGSAFATAVSSCSWCDSQ 240
Qy 241 APPEILCYLOSFWTGSFILANFDSRSRSGKDANTLLGSIHTFPDPAACDDSTFQPCSPRALA 300
```

```
Db 241 APPEILCYLOSFWTGSFILANFDSRSRSGKDANTLLGSIHTFPDPAACDDSTFQPCSPRALA 300
Qy 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRPEPTYNGNPMFLCTIAAEQYDALYQWD 360
Db 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRPEPTYNGNPMFLCTIAAEQYDALYQWD 360
Qy 361 KQGSLEVTVDLDFPFAALYSDAATGYSSSSSTYSSIVDAVKTFADGPFVSVETHAASNG 420
Db 361 KQGSLEVTVDLDFPFAALYSDAATGYSSSSSTYSSIVDAVKTFADGPFVSVETHAASNG 420
Qy 421 SMSEQYDKSDGDEQLSARDLTWGSYAALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
Db 421 SMSEQYDKSDGDEQLSARDLTWGSYAALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
Qy 481 TYSSVTVTWSPSIIVATGCTTTTATPTGSGSVTSTKTTATASKTSTTTTSGMSL 534
Db 481 TYSSVTVTWSPSIIVATGCTTTTATPTGSGSVTSTKTTATASKTSTTTTSGMSL 534

RESULT 4
US-10-421-586-2
; Sequence 2, Application US/10421586
; Publication No. US20040002142A1
; GENERAL INFORMATION:
; APPLICANT: Nielsen, Bjarne Roenfeldt
; APPLICANT: Svendsen, Allan
; APPLICANT: Pedersen, Henrik
; APPLICANT: Vind, Jesper
; APPLICANT: Hendriksen, Hanne Vang
; APPLICANT: Frandsen, Torben Peter
; TITLE OF INVENTION: Glucoamylase Variants
; FILE REFERENCE: 5667.210-US
; CURRENT APPLICATION NUMBER: US/10/421,586
; CURRENT FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-10-421-586-2

Query Match 100.0%; Score 2771; DB 15; Length 534;
Best Local Similarity 100.0%; Pred. No. 2,7e-237;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSFRSLALSGLVCTGLANVISKRATLDLSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
Db 1 MSFRSLALSGLVCTGLANVISKRATLDLSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
Qy 61 VVASPSTNDPDEYTYTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIVQGISNPSG 120
Db 61 VVASPSTNDPDEYTYTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIVQGISNPSG 120
Qy 121 DLSSGAGLEPKKNVDETAYTGSWGRPDGPALRATAMIGFQWLLDNGYSTTADIW 180
Db 121 DLSSGAGLEPKKNVDETAYTGSWGRPDGPALRATAMIGFQWLLDNGYSTTADIW 180
Qy 181 PLVRNDLSYVAQYWNQGTGYDLMEEVNGSFFTTIAVQHRALVEGSAFATAVSSCSWCDSQ 240
Db 181 PLVRNDLSYVAQYWNQGTGYDLMEEVNGSFFTTIAVQHRALVEGSAFATAVSSCSWCDSQ 240
Qy 241 APPEILCYLOSFWTGSFILANFDSRSRSGKDANTLLGSIHTFPDPAACDDSTFQPCSPRALA 300
Db 241 APPEILCYLOSFWTGSFILANFDSRSRSGKDANTLLGSIHTFPDPAACDDSTFQPCSPRALA 300
Qy 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRPEPTYNGNPMFLCTIAAEQYDALYQWD 360
Db 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRPEPTYNGNPMFLCTIAAEQYDALYQWD 360
Qy 361 KQGSLEVTVDLDFPFAALYSDAATGYSSSSSTYSSIVDAVKTFADGPFVSVETHAASNG 420
```

Db 361 KQSGLEVTDVSLDPFKALYSDAATGTYSSSSSTYSIVDAVKTFADGFSIVETHAASNG 420  
Qy 421 SMSEQYDKSDGEOLASARDLTWYSYALLTANNRRNSVVPASMGETSASSVGTCAATSAIG 480  
Db 421 SMSEQYDKSDGEOLASARDLTWYSYALLTANNRRNSVVPASMGETSASSVGTCAATSAIG 480  
Qy 481 TYSSVTVTSMPSIVATGCTTTTATPTGSGSVTSTKTTATATASKTSTTTRGMSL 534  
Db 481 TYSSVTVTSMPSIVATGCTTTTATPTGSGSVTSTKTTATATASKTSTTTRGMSL 534

RESULT 5  
US-10-421-586-3  
; Sequence 3, Application US/10421586  
; Publication No. US20040002142A1  
; GENERAL INFORMATION:  
; APPLICANT: Nielsen, Bjarne Ronfeldt  
; APPLICANT: Svendsen, Allan  
; APPLICANT: Pedersen, Henrik  
; APPLICANT: Vind, Jesper  
; APPLICANT: Hendriksen, Hanne Vang  
; APPLICANT: Frandsen, Torben Peter  
; TITLE OF INVENTION: Glucoamylase Variants  
; FILE REFERENCE: 5967 210-US  
; CURRENT APPLICATION NUMBER: US/10/421,586  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 3  
; LENGTH: 534  
; TYPE: PRF  
; ORGANISM: Aspergillus niger  
; FEATURE:  
; NAME/KEY: Signal  
; LOCATION: (1)..(24)  
US-10-421-586-3

Query Match 100.0%; Score 2771; DB 15; Length 534;  
Best Local Similarity 100.0%; Pred. No. 2,7e-237;  
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSFRSLIALSGVCTGLANYSKRATLDSWLSNEATYARFAIINNIGADGAWYSGADSGI 60  
Db 1 MSFRSLIALSGVCTGLANYSKRATLDSWLSNEATYARFAIINNIGADGAWYSGADSGI 60  
Qy 61 VVASPSTNDPDIYFTYTRDSGLVKTLYDLFRNGDTSLSSTIENYISAQAIVOGISNPSG 120  
Db 61 VVASPSTNDPDIYFTYTRDSGLVKTLYDLFRNGDTSLSSTIENYISAQAIVOGISNPSG 120  
Qy 121 DLSGGAGLGPKNVDETAATYTGSMGRPQRDPALRATAMIGFGQWLLDNGYSTATDIW 180  
Db 121 DLSGGAGLGPKNVDETAATYTGSMGRPQRDPALRATAMIGFGQWLLDNGYSTATDIW 180  
Qy 181 PLYRNDLSYVAQYWNQGYDLMEEVNGSSFFTTIAVQHRALVEGSAFATAVGSSCSWCDQ 240  
Db 181 PLYRNDLSYVAQYWNQGYDLMEEVNGSSFFTTIAVQHRALVEGSAFATAVGSSCSWCDQ 240  
Qy 241 APRILCYLQSFMTGSLFLANFDSRSKDNATLLGSIHTFDPBAACDSTFOPCSPRALA 300  
Db 241 APRILCYLQSFMTGSLFLANFDSRSKDNATLLGSIHTFDPBAACDSTFOPCSPRALA 300  
Qy 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDITYNGNPMFLCTLLAAEQLYDALYQMD 360  
Db 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDITYNGNPMFLCTLLAAEQLYDALYQMD 360  
Qy 361 KQSGLEVTDVSLDPFKALYSDAATGTYSSSSSTYSIVDAVKTFADGFSIVETHAASNG 420  
Db 361 KQSGLEVTDVSLDPFKALYSDAATGTYSSSSSTYSIVDAVKTFADGFSIVETHAASNG 420  
Qy 421 SMSEQYDKSDGEOLASARDLTWYSYALLTANNRRNSVVPASMGETSASSVGTCAATSAIG 480  
Db 421 SMSEQYDKSDGEOLASARDLTWYSYALLTANNRRNSVVPASMGETSASSVGTCAATSAIG 480

Qy 481 TYSSVTVTSMPSIVATGCTTTTATPTGSGSVTSTKTTATATASKTSTTTRGMSL 534  
Db 481 TYSSVTVTSMPSIVATGCTTTTATPTGSGSVTSTKTTATATASKTSTTTRGMSL 534

RESULT 6  
US-10-038-723-13  
; Sequence 13, Application US/10038723  
; Publication No. US20030032163A1  
; GENERAL INFORMATION:  
; APPLICANT: Nielsen, Bjarne Ronfeldt  
; APPLICANT: Svendsen, Allan  
; APPLICANT: Pedersen, Henrik  
; APPLICANT: Vind, Jesper  
; APPLICANT: Hendriksen, Hanne Vang  
; APPLICANT: Frandsen, Torben Peter  
; TITLE OF INVENTION: Glucoamylase Variants  
; FILE REFERENCE: 5636,200-US  
; CURRENT APPLICATION NUMBER: US/10/038,723  
; CURRENT FILING DATE: 2002-01-02  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/351,814  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-12  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PA 1998 00937  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-15  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PA 1998 01667  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-12-17  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/093,528  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-21  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/115,545  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-12  
; NUMBER OF SEQ ID NOS: 81  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 13  
; LENGTH: 640  
; TYPE: PRF  
; ORGANISM: ASPERGILLUS NIGER  
US-10-038-723-13

Query Match 99.0%; Score 2742; DB 14; Length 640;  
Best Local Similarity 99.6%; Pred. No. 1,4e-234;  
Matches 528; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSFRSLIALSGVCTGLANYSKRATLDSWLSNEATYARFAIINNIGADGAWYSGADSGI 60  
Db 1 MSFRSLIALSGVCTGLANYSKRATLDSWLSNEATYARFAIINNIGADGAWYSGADSGI 60  
Qy 61 VVASPSTNDPDIYFTYTRDSGLVKTLYDLFRNGDTSLSSTIENYISAQAIVOGISNPSG 120  
Db 61 VVASPSTNDPDIYFTYTRDSGLVKTLYDLFRNGDTSLSSTIENYISAQAIVOGISNPSG 120  
Qy 121 DLSGGAGLGPKNVDETAATYTGSMGRPQRDPALRATAMIGFGQWLLDNGYSTATDIW 180  
Db 121 DLSGGAGLGPKNVDETAATYTGSMGRPQRDPALRATAMIGFGQWLLDNGYSTATDIW 180  
Qy 181 PLYRNDLSYVAQYWNQGYDLMEEVNGSSFFTTIAVQHRALVEGSAFATAVGSSCSWCDQ 240  
Db 181 PLYRNDLSYVAQYWNQGYDLMEEVNGSSFFTTIAVQHRALVEGSAFATAVGSSCSWCDQ 240  
Qy 241 APRILCYLQSFMTGSLFLANFDSRSKDNATLLGSIHTFDPBAACDSTFOPCSPRALA 300  
Db 241 APRILCYLQSFMTGSLFLANFDSRSKDNATLLGSIHTFDPBAACDSTFOPCSPRALA 300  
Qy 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDITYNGNPMFLCTLLAAEQLYDALYQMD 360  
Db 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDITYNGNPMFLCTLLAAEQLYDALYQMD 360  
Qy 361 KQSGLEVTDVSLDPFKALYSDAATGTYSSSSSTYSIVDAVKTFADGFSIVETHAASNG 420  
Db 361 KQSGLEVTDVSLDPFKALYSDAATGTYSSSSSTYSIVDAVKTFADGFSIVETHAASNG 420  
Qy 421 SMSEQYDKSDGEOLASARDLTWYSYALLTANNRRNSVVPASMGETSASSVGTCAATSAIG 480  
Db 421 SMSEQYDKSDGEOLASARDLTWYSYALLTANNRRNSVVPASMGETSASSVGTCAATSAIG 480



```

; GENERAL INFORMATION:
; APPLICANT: Power, Scott D.
; APPLICANT: Wang, Huaming
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Production of Functional Antibodies in
; TITLE OF INVENTION: Filamentous Fungi
; FILE REFERENCE: GC741-2
; CURRENT APPLICATION NUMBER: US/10/418,836
; CURRENT FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US 60/373,889
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/411,540
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/452,134
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 60/411,537
; PRIOR FILING DATE: 2002-09-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 763
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion protein
US-10-418-836-13

```

Query Match 98.7%; Score 2735; DB 15; Length 763;  
Best Local Similarity 99.6%; Pred. No. 7.5e-234;  
Matches 527; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 MSRRSLALSGVCTGLANVISKRATLDSWLSNEATVAATKALININIGADGAWVSGADSGI 60
DB 1 MSRRSLALSGVCTGLANVISKRATLDSWLSNEATVAATKALININIGADGAWVSGADSGI 60
QY 61 VVASPTDNDPYYTTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIYOGISNPSG 120
DB 61 VVASPTDNDPYYTTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIYOGISNPSG 120
QY 121 DLSGAGLGEPRKNNVDETAATYTGSGRQORDGAPALRATAMIGFGQWLLDNGYSTATDIW 180
DB 121 DLSGAGLGEPRKNNVDETAATYTGSGRQORDGAPALRATAMIGFGQWLLDNGYSTATDIW 180
QY 121 DLSGAGLGEPRKNNVDETAATYTGSGRQORDGAPALRATAMIGFGQWLLDNGYSTATDIW 180
DB 121 DLSGAGLGEPRKNNVDETAATYTGSGRQORDGAPALRATAMIGFGQWLLDNGYSTATDIW 180
QY 181 PLVRNDLSYVAQYWNQTYDLMEEVNNGSSFFTTIAVQHRALVBSAFAATVAGSSCWCDSQ 240
DB 181 PLVRNDLSYVAQYWNQTYDLMEEVNNGSSFFTTIAVQHRALVBSAFAATVAGSSCWCDSQ 240
QY 241 APRILCYLGSFMTGSFTLANFDSRSRSGKDANTLLGSIHTFDPPEAACDDSTFQPCSPRALA 300
DB 241 APRILCYLGSFMTGSFTLANFDSRSRSGKDANTLLGSIHTFDPPEAACDDSTFQPCSPRALA 300
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLLAAAEQLYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLLAAAEQLYDALYQMD 360
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLLAAAEQLYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLLAAAEQLYDALYQMD 360
QY 361 KQSLSEVTIVSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTPADGPFVSIYETHAASNG 420
DB 361 KQSLSEVTIVSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTPADGPFVSIYETHAASNG 420
QY 361 KQSLSEVTIVSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTPADGPFVSIYETHAASNG 420
DB 361 KQSLSEVTIVSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTPADGPFVSIYETHAASNG 420
QY 421 SMSEQYDKSDGEOLSARDLTWVSFAALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGEOLSARDLTWVSFAALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
QY 421 SMSEQYDKSDGEOLSARDLTWVSFAALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGEOLSARDLTWVSFAALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
QY 481 TVSSVTVTSPSIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTYKR 529
DB 481 TVSSVTVTSPSIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTYKR 529

```

RESULT 10  
US-10-418-836-10  
; Sequence 10, Application US/10418836  
; Publication No. US20040018573A1  
; GENERAL INFORMATION:  
; APPLICANT: Power, Scott D.

```

; APPLICANT: Power, Scott D.
; APPLICANT: Wang, Huaming
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Production of Functional Antibodies in
; TITLE OF INVENTION: Filamentous Fungi
; FILE REFERENCE: GC741-2
; CURRENT APPLICATION NUMBER: US/10/418,836
; CURRENT FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US 60/373,889
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/411,540
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/452,134
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 60/411,537
; PRIOR FILING DATE: 2002-09-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 979
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion protein
US-10-418-836-10

```

Query Match 98.7%; Score 2735; DB 15; Length 979;  
Best Local Similarity 99.6%; Pred. No. 1.1e-233;  
Matches 527; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 MSRRSLALSGVCTGLANVISKRATLDSWLSNEATVAATKALININIGADGAWVSGADSGI 60
DB 1 MSRRSLALSGVCTGLANVISKRATLDSWLSNEATVAATKALININIGADGAWVSGADSGI 60
QY 61 VVASPTDNDPYYTTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIYOGISNPSG 120
DB 61 VVASPTDNDPYYTTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIYOGISNPSG 120
QY 121 DLSGAGLGEPRKNNVDETAATYTGSGRQORDGAPALRATAMIGFGQWLLDNGYSTATDIW 180
DB 121 DLSGAGLGEPRKNNVDETAATYTGSGRQORDGAPALRATAMIGFGQWLLDNGYSTATDIW 180
QY 121 DLSGAGLGEPRKNNVDETAATYTGSGRQORDGAPALRATAMIGFGQWLLDNGYSTATDIW 180
DB 121 DLSGAGLGEPRKNNVDETAATYTGSGRQORDGAPALRATAMIGFGQWLLDNGYSTATDIW 180
QY 181 PLVRNDLSYVAQYWNQTYDLMEEVNNGSSFFTTIAVQHRALVBSAFAATVAGSSCWCDSQ 240
DB 181 PLVRNDLSYVAQYWNQTYDLMEEVNNGSSFFTTIAVQHRALVBSAFAATVAGSSCWCDSQ 240
QY 241 APRILCYLGSFMTGSFTLANFDSRSRSGKDANTLLGSIHTFDPPEAACDDSTFQPCSPRALA 300
DB 241 APRILCYLGSFMTGSFTLANFDSRSRSGKDANTLLGSIHTFDPPEAACDDSTFQPCSPRALA 300
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLLAAAEQLYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLLAAAEQLYDALYQMD 360
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLLAAAEQLYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLLAAAEQLYDALYQMD 360
QY 361 KQSLSEVTIVSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTPADGPFVSIYETHAASNG 420
DB 361 KQSLSEVTIVSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTPADGPFVSIYETHAASNG 420
QY 361 KQSLSEVTIVSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTPADGPFVSIYETHAASNG 420
DB 361 KQSLSEVTIVSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTPADGPFVSIYETHAASNG 420
QY 421 SMSEQYDKSDGEOLSARDLTWVSFAALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGEOLSARDLTWVSFAALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
QY 421 SMSEQYDKSDGEOLSARDLTWVSFAALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGEOLSARDLTWVSFAALLTANNRNSVVPASWGETSASVPGTCAATSAIG 480
QY 481 TVSSVTVTSPSIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTYKR 529
DB 481 TVSSVTVTSPSIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTYKR 529

```

RESULT 11  
US-10-418-836-16  
; Sequence 16, Application US/10418836  
; Publication No. US20040018573A1  
; GENERAL INFORMATION:  
; APPLICANT: Power, Scott D.

APPLICANT: Wang, Huaming  
APPLICANT: Ward, Michael  
TITLE OF INVENTION: Production of Functional Antibodies in  
FILE REFERENCE: GC741-2  
CURRENT APPLICATION NUMBER: US/10/418, 836  
CURRENT FILING DATE: 2003-04-17  
PRIOR APPLICATION NUMBER: US 60/373, 889  
PRIOR FILING DATE: 2002-04-18  
PRIOR APPLICATION NUMBER: US 60/411, 540  
PRIOR FILING DATE: 2002-09-18  
PRIOR APPLICATION NUMBER: US 60/452, 134  
PRIOR FILING DATE: 2003-03-04  
PRIOR APPLICATION NUMBER: US 60/411, 537  
PRIOR FILING DATE: 2002-09-18  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 16  
LENGTH: 979  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: fusion protein  
US-10-418-836-16

Query Match 98.7%; Score 2735; DB 15; Length 979;  
Best Local Similarity 99.6%; Pred. No. 1,1e-233;  
Matches 527; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSFRSLALSGVCTGLANVISKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60  
DB 1 MSFRSLALSGVCTGLANVISKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60  
QY 61 VVASPSTNDPDIYTYTRDSGLVLTVDLFRNGDTSLSLTENYISAQAIYVGISNPSG 120  
DB 61 VVASPSTNDPDIYTYTRDSGLVLTVDLFRNGDTSLSLTENYISAQAIYVGISNPSG 120  
QY 121 DLSSGAGLGEPEKNVDETAYTGSWGRPQDGPALRATAMIGFGOMLLDNGYSTATDIYV 180  
DB 121 DLSSGAGLGEPEKNVDETAYTGSWGRPQDGPALRATAMIGFGOMLLDNGYSTATDIYV 180  
QY 181 PLVRNDLSYVAQYWNQGYDLMEVNGSSPFTIYVQHRALVEGSAFAVAVGSSCWCDQ 240  
DB 181 PLVRNDLSYVAQYWNQGYDLMEVNGSSPFTIYVQHRALVEGSAFAVAVGSSCWCDQ 240  
QY 241 APEILCYLOSFWTGSFTLANFDSRSRSGKANTLIGSIHTFDPPEAACDDSTFOCSPRALA 300  
DB 241 APEILCYLOSFWTGSFTLANFDSRSRSGKANTLIGSIHTFDPPEAACDDSTFOCSPRALA 300  
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTTAAAEQLYDALYQMD 360  
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTTAAAEQLYDALYQMD 360  
QY 361 KQGLEVTVDVSLDFPKALYSDAATGYSSSSSTYSSIVDAVKTFADGFVSI VETHAASNG 420  
DB 361 KQGLEVTVDVSLDFPKALYSDAATGYSSSSSTYSSIVDAVKTFADGFVSI VETHAASNG 420  
QY 421 SMEQYDKSDGEBLSARDLWTSYALLTANNRNSVVPASWGTSSASVPGTCAATSAG 480  
DB 421 SMEQYDKSDGEBLSARDLWTSYALLTANNRNSVVPASWGTSSASVPGTCAATSAG 480  
QY 481 TYSSTVTVTSPSIIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTYKR 529  
DB 481 TYSSTVTVTSPSIIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTYKR 529

RESULT 12  
US-10-418-836-19  
; Sequence 19, Application US/10418836  
; Publication No. US20040018573A1  
; GENERAL INFORMATION:  
; APPLICANT: Power, Scott D.  
; APPLICANT: Wang, Huaming

APPLICANT: Ward, Michael  
APPLICANT: Michael  
TITLE OF INVENTION: Production of Functional Antibodies in  
FILE REFERENCE: GC741-2  
CURRENT APPLICATION NUMBER: US/10/418, 836  
CURRENT FILING DATE: 2003-04-17  
PRIOR APPLICATION NUMBER: US 60/373, 889  
PRIOR FILING DATE: 2002-04-18  
PRIOR APPLICATION NUMBER: US 60/411, 540  
PRIOR FILING DATE: 2002-09-18  
PRIOR APPLICATION NUMBER: US 60/452, 134  
PRIOR FILING DATE: 2003-03-04  
PRIOR APPLICATION NUMBER: US 60/411, 537  
PRIOR FILING DATE: 2002-09-18  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 19  
LENGTH: 738  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: fusion protein  
US-10-418-836-19

Query Match 98.2%; Score 2720.5; DB 15; Length 738;  
Best Local Similarity 98.5%; Pred. No. 1,4e-232;  
Matches 527; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 1 MSFRSLALSGVCTGLANVISKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60  
DB 1 MSFRSLALSGVCTGLANVISKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60  
QY 61 VVASPSTNDPDIYTYTRDSGLVLTVDLFRNGDTSLSLTENYISAQAIYVGISNPSG 120  
DB 61 VVASPSTNDPDIYTYTRDSGLVLTVDLFRNGDTSLSLTENYISAQAIYVGISNPSG 120  
QY 121 DLSSGAGLGEPEKNVDETAYTGSWGRPQDGPALRATAMIGFGOMLLDNGYSTATDIYV 180  
DB 121 DLSSGAGLGEPEKNVDETAYTGSWGRPQDGPALRATAMIGFGOMLLDNGYSTATDIYV 180  
QY 181 PLVRNDLSYVAQYWNQGYDLMEVNGSSPFTIYVQHRALVEGSAFAVAVGSSCWCDQ 240  
DB 181 PLVRNDLSYVAQYWNQGYDLMEVNGSSPFTIYVQHRALVEGSAFAVAVGSSCWCDQ 240  
QY 241 APEILCYLOSFWTGSFTLANFDSRSRSGKANTLIGSIHTFDPPEAACDDSTFOCSPRALA 300  
DB 241 APEILCYLOSFWTGSFTLANFDSRSRSGKANTLIGSIHTFDPPEAACDDSTFOCSPRALA 300  
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTTAAAEQLYDALYQMD 360  
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTTAAAEQLYDALYQMD 360  
QY 361 KQGLEVTVDVSLDFPKALYSDAATGYSSSSSTYSSIVDAVKTFADGFVSI VETHAASNG 420  
DB 361 KQGLEVTVDVSLDFPKALYSDAATGYSSSSSTYSSIVDAVKTFADGFVSI VETHAASNG 420  
QY 421 SMEQYDKSDGEBLSARDLWTSYALLTANNRNSVVPASWGTSSASVPGTCAATSAG 480  
DB 421 SMEQYDKSDGEBLSARDLWTSYALLTANNRNSVVPASWGTSSASVPGTCAATSAG 480  
QY 481 TYSSTVTVTSPSIIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTYKR 534  
DB 481 TYSSTVTVTSPSIIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTYKR 534

RESULT 13  
US-10-418-836-30  
; Sequence 30, Application US/10418836  
; Publication No. US20040018573A1  
; GENERAL INFORMATION:  
; APPLICANT: Power, Scott D.  
; APPLICANT: Wang, Huaming  
; APPLICANT: Ward, Michael

```
; TITLE OF INVENTION: Production of Functional Antibodies in
; FILE REFERENCE: GC741-2
; CURRENT APPLICATION NUMBER: US/10/418,836
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US 60/373,889
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/411,540
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/452,134
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 60/411,537
; PRIOR FILING DATE: 2002-09-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion protein
US-10-418-836-30
```

```
Query Match      98.0%; Score 2716; DB 15; Length 741;
Best Local Similarity 98.0%; Pred. No. 3.5e-232;
Matches 527; Conservative 1; Mismatches 6; Indels 4; Gaps 1;
```

```
QY      1 MSFRLSLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
Db      1 MSFRLSLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
QY      61 VVASPSTDNPDYFYTWTRDSGLVLTIVLDFRNGDTSLSLTINYSIAQAIVGGINSPSG 120
Db      61 VVASPSTDNPDYFYTWTRDSGLVLTIVLDFRNGDTSLSLTINYSIAQAIVGGINSPSG 120
QY      121 DLSGAGLGEKPKFNVDEATYTGSMGRPQRDGPALRAMIGFGQWLLDNGYTSATDIW 180
Db      121 DLSGAGLGEKPKFNVDEATYTGSMGRPQRDGPALRAMIGFGQWLLDNGYTSATDIW 180
QY      181 PLVNDLSYVAQYWNQGYDLMEEVNGSSFFTIAVOHRAIVEGSAFATAVSSCSMCDSQ 240
Db      181 PLVNDLSYVAQYWNQGYDLMEEVNGSSFFTIAVOHRAIVEGSAFATAVSSCSMCDSQ 240
QY      241 APEILCYLQSFMTGSFTLANFDSRSRGKANTLLGSIHTFDPEAACDDSTFQPCSPRALA 300
Db      241 APEILCYLQSFMTGSFTLANFDSRSRGKANTLLGSIHTFDPEAACDDSTFQPCSPRALA 300
QY      301 NHRKVVDSPRSITYLNDGLSDSEAVAGRYPEDTYNGNPFVLTCTLAABQLYDALYQMD 360
Db      301 NHRKVVDSPRSITYLNDGLSDSEAVAGRYPEDTYNGNPFVLTCTLAABQLYDALYQMD 360
QY      361 KQGSLEVTDVSLDFFKALYSDAATGTYSSSSSTYSIVDAVKTFAFGFVSIIVETHAASNG 420
Db      361 KQGSLEVTDVSLDFFKALYSDAATGTYSSSSSTYSIVDAVKTFAFGFVSIIVETHAASNG 420
QY      421 SMSBQYDKSDGEQLSARDLTWSYAALLTANNRRNSVVPASWGTSASVPGTCAATSAIG 480
Db      421 SMSBQYDKSDGEQLSARDLTWSYAALLTANNRRNSVVPASWGTSASVPGTCAATSAIG 480
QY      481 TYSSVVTWSPSIYVAGCGTTTATPTGSGSVTSKTTATASKTS-----TTRSGMSL 534
Db      481 TYSSVVTWSPSIYVAGCGTTTATPTGSGSVTSKTTATASKTS-----TTRSGMSL 534
```

```
RESULT 14
US-10-418-836-38
; Sequence 38, Application US/10418836
; Publication No. US20040018573A1
; GENERAL INFORMATION:
; APPLICANT: Power, Scott D.
; APPLICANT: Wang, Huaming
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Production of Functional Antibodies in
```

```
; TITLE OF INVENTION: Filamentous Fungi
; FILE REFERENCE: GC741-2
; CURRENT APPLICATION NUMBER: US/10/418,836
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US 60/373,889
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/411,540
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/452,134
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 60/411,537
; PRIOR FILING DATE: 2002-09-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 972
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion protein
US-10-418-836-38
```

```
Query Match      98.0%; Score 2715.5; DB 15; Length 972;
Best Local Similarity 95.5%; Pred. No. 5.8e-232;
Matches 530; Conservative 0; Mismatches 4; Indels 21; Gaps 2;
```

```
QY      1 MSFRLSLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
Db      1 MSFRLSLALSGLVCTGLANVISKRATLDSWLSNEATVARTAILNNIGADGAWVSGADSGI 60
QY      61 VVASPSTDNPDYFYTWTRDSGLVLTIVLDFRNGDTSLSLTINYSIAQAIVGGINSPSG 120
Db      61 VVASPSTDNPDYFYTWTRDSGLVLTIVLDFRNGDTSLSLTINYSIAQAIVGGINSPSG 120
QY      121 DLSGAGLGEKPKFNVDEATYTGSMGRPQRDGPALRAMIGFGQWLLDNGYTSATDIW 180
Db      121 DLSGAGLGEKPKFNVDEATYTGSMGRPQRDGPALRAMIGFGQWLLDNGYTSATDIW 180
QY      181 PLVNDLSYVAQYWNQGYDLMEEVNGSSFFTIAVOHRAIVEGSAFATAVSSCSMCDSQ 240
Db      181 PLVNDLSYVAQYWNQGYDLMEEVNGSSFFTIAVOHRAIVEGSAFATAVSSCSMCDSQ 240
QY      241 APEILCYLQSFMTGSFTLANFDSRSRGKANTLLGSIHTFDPEAACDDSTFQPCSPRALA 300
Db      241 APEILCYLQSFMTGSFTLANFDSRSRGKANTLLGSIHTFDPEAACDDSTFQPCSPRALA 300
QY      301 NHRKVVDSPRSITYLNDGLSDSEAVAGRYPEDTYNGNPFVLTCTLAABQLYDALYQMD 360
Db      301 NHRKVVDSPRSITYLNDGLSDSEAVAGRYPEDTYNGNPFVLTCTLAABQLYDALYQMD 360
QY      361 KQGSLEVTDVSLDFFKALYSDAATGTYSSSSSTYSIVDAVKTFAFGFVSIIVETHAASNG 420
Db      361 KQGSLEVTDVSLDFFKALYSDAATGTYSSSSSTYSIVDAVKTFAFGFVSIIVETHAASNG 420
QY      421 SMSBQYDKSDGEQLSARDLTWSYAALLTANNRRNSVVPASWGTSASVPGTCAATSAIG 480
Db      421 SMSBQYDKSDGEQLSARDLTWSYAALLTANNRRNSVVPASWGTSASVPGTCAATSAIG 480
QY      481 TYSSVVTWSPSIYVAGCGTTTATPTGSGSVTSKTTATASKTS-----TTRSGMSL 534
Db      481 TYSSVVTWSPSIYVAGCGTTTATPTGSGSVTSKTTATASKTS-----TTRSGMSL 534
```

```
RESULT 15
US-10-418-836-26
; Sequence 26, Application US/10418836
; Publication No. US20040018573A1
; GENERAL INFORMATION:
; APPLICANT: Power, Scott D.
```



APPLICANT: Wang, Huaming  
APPLICANT: Ward, Michael  
TITLE OF INVENTION: Production of Functional Antibodies in  
TITLE OF INVENTION: Filamentous Fungi  
FILE REFERENCE: GC741-2  
CURRENT APPLICATION NUMBER: US/10/418,836  
CURRENT FILING DATE: 2003-04-17  
PRIOR APPLICATION NUMBER: US 60/373,889  
PRIOR FILING DATE: 2002-04-18  
PRIOR APPLICATION NUMBER: US 60/411,540  
PRIOR FILING DATE: 2002-09-18  
PRIOR APPLICATION NUMBER: US 60/452,134  
PRIOR FILING DATE: 2003-03-04  
PRIOR APPLICATION NUMBER: US 60/411,537  
PRIOR FILING DATE: 2002-09-18  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: FaastSeq for Windows Version 4.0  
SEQ ID NO 26  
LENGTH: 739  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: fusion protein  
US-10-418-836-26

Query Match 98.0%; Score 2715; DB 15; Length 739;  
Best Local Similarity 99.8%; Pred. No. 4.3e-232;  
Matches 522; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60  
DB 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60  
QY 61 VVASPTDNDPDIYFYWTRDSGLVLTVDLFRNGDTSLSLTENYISAQAIYVGISNPSG 120  
DB 61 VVASPTDNDPDIYFYWTRDSGLVLTVDLFRNGDTSLSLTENYISAQAIYVGISNPSG 120  
QY 121 DLSGAGLGEPEKFNVDIYVTSWGRPORDGALRATAMIGFQWLLDNGYSTATDIYW 180  
DB 121 DLSGAGLGEPEKFNVDIYVTSWGRPORDGALRATAMIGFQWLLDNGYSTATDIYW 180  
QY 181 PLVRNDLSYVAQYWNQGYDLMEEVNGSSFTTIAVQHRALVEGSAFATVAGSSCWCDG 240  
DB 181 PLVRNDLSYVAQYWNQGYDLMEEVNGSSFTTIAVQHRALVEGSAFATVAGSSCWCDG 240  
QY 241 APEILCYLOSFWTGSFTILANFDSRSRSGKDNLTLSITHTDPEAACDDSTFPQCSPRALA 300  
DB 241 APEILCYLOSFWTGSFTILANFDSRSRSGKDNLTLSITHTDPEAACDDSTFPQCSPRALA 300  
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAALAEQLYDALYQMD 360  
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAALAEQLYDALYQMD 360  
QY 361 KQGLEVTDVSLDFPRLALYSDAATGTYSSSSSTYSIVDAVKTFFADGFSIVETHAASNG 420  
DB 361 KQGLEVTDVSLDFPRLALYSDAATGTYSSSSSTYSIVDAVKTFFADGFSIVETHAASNG 420  
QY 421 SMSEQYDKSDGEOLSARDLTWSYAALLTANNRNSVVPASWGETSASSVPGTCAATSAIG 480  
DB 421 SMSEQYDKSDGEOLSARDLTWSYAALLTANNRNSVVPASWGETSASSVPGTCAATSAIG 480  
QY 481 TYSSVVTWSPSIVATGCTTTTATPTGSGSVTSKTTATASK 523  
DB 481 TYSSVVTWSPSIVATGCTTTTATPTGSGSVTSKTTATASK 523

RESULT 16  
US-10-418-836-27  
Sequence 27, Application US/10418836  
Publication No. US20040018573A1  
GENERAL INFORMATION:  
APPLICANT: Power, Scott D.  
APPLICANT: Wang, Huaming

APPLICANT: Ward, Michael  
APPLICANT: Ward, Michael  
TITLE OF INVENTION: Production of Functional Antibodies in  
TITLE OF INVENTION: Filamentous Fungi  
FILE REFERENCE: GC741-2  
CURRENT APPLICATION NUMBER: US/10/418,836  
CURRENT FILING DATE: 2003-04-17  
PRIOR APPLICATION NUMBER: US 60/373,889  
PRIOR FILING DATE: 2002-04-18  
PRIOR APPLICATION NUMBER: US 60/411,540  
PRIOR FILING DATE: 2002-09-18  
PRIOR APPLICATION NUMBER: US 60/452,134  
PRIOR FILING DATE: 2003-03-04  
PRIOR APPLICATION NUMBER: US 60/411,537  
PRIOR FILING DATE: 2002-09-18  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: FaastSeq for Windows Version 4.0  
SEQ ID NO 27  
LENGTH: 739  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: fusion protein  
US-10-418-836-27

Query Match 98.0%; Score 2715; DB 15; Length 739;  
Best Local Similarity 98.1%; Pred. No. 4.3e-232;  
Matches 526; Conservative 2; Mismatches 6; Indels 2; Gaps 1;

QY 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60  
DB 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60  
QY 61 VVASPTDNDPDIYFYWTRDSGLVLTVDLFRNGDTSLSLTENYISAQAIYVGISNPSG 120  
DB 61 VVASPTDNDPDIYFYWTRDSGLVLTVDLFRNGDTSLSLTENYISAQAIYVGISNPSG 120  
QY 121 DLSGAGLGEPEKFNVDIYVTSWGRPORDGALRATAMIGFQWLLDNGYSTATDIYW 180  
DB 121 DLSGAGLGEPEKFNVDIYVTSWGRPORDGALRATAMIGFQWLLDNGYSTATDIYW 180  
QY 181 PLVRNDLSYVAQYWNQGYDLMEEVNGSSFTTIAVQHRALVEGSAFATVAGSSCWCDG 240  
DB 181 PLVRNDLSYVAQYWNQGYDLMEEVNGSSFTTIAVQHRALVEGSAFATVAGSSCWCDG 240  
QY 241 APEILCYLOSFWTGSFTILANFDSRSRSGKDNLTLSITHTDPEAACDDSTFPQCSPRALA 300  
DB 241 APEILCYLOSFWTGSFTILANFDSRSRSGKDNLTLSITHTDPEAACDDSTFPQCSPRALA 300  
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAALAEQLYDALYQMD 360  
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAALAEQLYDALYQMD 360  
QY 361 KQGLEVTDVSLDFPRLALYSDAATGTYSSSSSTYSIVDAVKTFFADGFSIVETHAASNG 420  
DB 361 KQGLEVTDVSLDFPRLALYSDAATGTYSSSSSTYSIVDAVKTFFADGFSIVETHAASNG 420  
QY 421 SMSEQYDKSDGEOLSARDLTWSYAALLTANNRNSVVPASWGETSASSVPGTCAATSAIG 480  
DB 421 SMSEQYDKSDGEOLSARDLTWSYAALLTANNRNSVVPASWGETSASSVPGTCAATSAIG 480  
QY 481 TYSSVVTWSPSIVATGCTTTTATPTGSGSVTSKTTATASK -KTSYTRSGMSL 534  
DB 481 TYSSVVTWSPSIVATGCTTTTATPTGSGSVTSKTTATASKKRDIOQTGSSSL 536

RESULT 17  
US-10-418-836-28  
Sequence 28, Application US/10418836  
Publication No. US20040018573A1  
GENERAL INFORMATION:  
APPLICANT: Power, Scott D.  
APPLICANT: Wang, Huaming  
APPLICANT: Ward, Michael

```
; TITLE OF INVENTION: Production of Functional Antibodies in
; FILE REFERENCE: GCT41-2
; CURRENT APPLICATION NUMBER: US/10/418,836
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US 60/373,889
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/411,540
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/452,134
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 60/411,537
; PRIOR FILING DATE: 2002-09-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion protein
US-10-418-836-28
```

```
Query Match          98.0%; Score 2714.5; DB 15; Length 740;
Best Local Similarity 98.0%; Pred. No. 4,7e-232;
Matches 526; Conservative 2; Mismatches 6; Indels 3; Gaps 1;
```

```
QY 1 MSFRLSLALSGLVCTGLANVISKRATLDSWLSNEATVARTAIINNIGADGAWVSGADSGI 60
DB 1 MSFRLSLALSGLVCTGLANVISKRATLDSWLSNEATVARTAIINNIGADGAWVSGADSGI 60
QY 61 VVASPSTNDPDIYTYWTRDSGLVLTIVLDFRNGDTSLSSTENYISAQAIIVGGINSPSG 120
DB 61 VVASPSTNDPDIYTYWTRDSGLVLTIVLDFRNGDTSLSSTENYISAQAIIVGGINSPSG 120
QY 121 DLSGAGLGEPRKFNVDATAITGSGWRPQRDGPALRATAMIGFGQWLLDNGYSTATDIYW 180
DB 121 DLSGAGLGEPRKFNVDATAITGSGWRPQRDGPALRATAMIGFGQWLLDNGYSTATDIYW 180
QY 181 PLVRNDLSYVAQYWNQGYDLMEEVNGSSFFTIAVQHRALVEGSAFATAVSSCSWCDSQ 240
DB 181 PLVRNDLSYVAQYWNQGYDLMEEVNGSSFFTIAVQHRALVEGSAFATAVSSCSWCDSQ 240
QY 241 APEILCYLGSFMTGSIILANFDSRSRSGKANTLLGSIHTFDPEAACDDSTFOPCSPRALA 300
DB 241 APEILCYLGSFMTGSIILANFDSRSRSGKANTLLGSIHTFDPEAACDDSTFOPCSPRALA 300
QY 301 NHKEVDSFRSITTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLLAAEQLDALYQMD 360
DB 301 NHKEVDSFRSITTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLLAAEQLDALYQMD 360
QY 361 KQGLSLEVTDVSLDFPKALYSDAATGTYSSTYSIVDAVKTFADGFVSIIVETHAASNG 420
DB 361 KQGLSLEVTDVSLDFPKALYSDAATGTYSSTYSIVDAVKTFADGFVSIIVETHAASNG 420
QY 421 SMSEQYDKSDGEOLASARDLTWSYAALLITANNRNSVVPASWGTSASVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGEOLASARDLTWSYAALLITANNRNSVVPASWGTSASVPGTCAATSAIG 480
QY 481 TYSSVVTWSPSIIVAGGTTTATPTGSGSVTSTSKTTATAS--KTSTTRSGMSL 534
DB 481 TYSSVVTWSPSIIVAGGTTTATPTGSGSVTSTSKTTATASISKRDIDMTQSPSSL 537
```

```
RESULT 18
US-10-418-836-34
; Sequence 34, Application US/10418836
; Publication No. US20040018573A1
; GENERAL INFORMATION:
; APPLICANT: Power, Scott D.
; APPLICANT: Wang, Huang
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Production of Functional Antibodies in
```

```
; TITLE OF INVENTION: Filamentous Fungi
; FILE REFERENCE: GCT41-2
; CURRENT APPLICATION NUMBER: US/10/418,836
; PRIOR FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US 60/373,889
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/411,540
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/452,134
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: US 60/411,537
; PRIOR FILING DATE: 2002-09-18
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fusion protein
US-10-418-836-34
```

```
Query Match          98.0%; Score 2714.5; DB 15; Length 740;
Best Local Similarity 98.0%; Pred. No. 4,7e-232;
Matches 526; Conservative 2; Mismatches 6; Indels 3; Gaps 1;
```

```
QY 1 MSFRLSLALSGLVCTGLANVISKRATLDSWLSNEATVARTAIINNIGADGAWVSGADSGI 60
DB 1 MSFRLSLALSGLVCTGLANVISKRATLDSWLSNEATVARTAIINNIGADGAWVSGADSGI 60
QY 61 VVASPSTNDPDIYTYWTRDSGLVLTIVLDFRNGDTSLSSTENYISAQAIIVGGINSPSG 120
DB 61 VVASPSTNDPDIYTYWTRDSGLVLTIVLDFRNGDTSLSSTENYISAQAIIVGGINSPSG 120
QY 121 DLSGAGLGEPRKFNVDATAITGSGWRPQRDGPALRATAMIGFGQWLLDNGYSTATDIYW 180
DB 121 DLSGAGLGEPRKFNVDATAITGSGWRPQRDGPALRATAMIGFGQWLLDNGYSTATDIYW 180
QY 181 PLVRNDLSYVAQYWNQGYDLMEEVNGSSFFTIAVQHRALVEGSAFATAVSSCSWCDSQ 240
DB 181 PLVRNDLSYVAQYWNQGYDLMEEVNGSSFFTIAVQHRALVEGSAFATAVSSCSWCDSQ 240
QY 241 APEILCYLGSFMTGSIILANFDSRSRSGKANTLLGSIHTFDPEAACDDSTFOPCSPRALA 300
DB 241 APEILCYLGSFMTGSIILANFDSRSRSGKANTLLGSIHTFDPEAACDDSTFOPCSPRALA 300
QY 301 NHKEVDSFRSITTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLLAAEQLDALYQMD 360
DB 301 NHKEVDSFRSITTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLLAAEQLDALYQMD 360
QY 361 KQGLSLEVTDVSLDFPKALYSDAATGTYSSTYSIVDAVKTFADGFVSIIVETHAASNG 420
DB 361 KQGLSLEVTDVSLDFPKALYSDAATGTYSSTYSIVDAVKTFADGFVSIIVETHAASNG 420
QY 421 SMSEQYDKSDGEOLASARDLTWSYAALLITANNRNSVVPASWGTSASVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGEOLASARDLTWSYAALLITANNRNSVVPASWGTSASVPGTCAATSAIG 480
QY 481 TYSSVVTWSPSIIVAGGTTTATPTGSGSVTSTSKTTATAS--KTSTTRSGMSL 534
DB 481 TYSSVVTWSPSIIVAGGTTTATPTGSGSVTSTSKTTATASISKRDIDMTQSPSSL 537
```

```
RESULT 19
US-10-418-836-29
; Sequence 29, Application US/10418836
; Publication No. US20040018573A1
; GENERAL INFORMATION:
; APPLICANT: Power, Scott D.
; APPLICANT: Wang, Huang
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Filamentous Fungi
```



FILE REFERENCE: GC741-2  
CURRENT APPLICATION NUMBER: US/10/418,836  
CURRENT FILING DATE: 2003-04-17  
PRIOR APPLICATION NUMBER: US 60/373,889  
PRIOR FILING DATE: 2002-04-18  
PRIOR APPLICATION NUMBER: US 60/411,540  
PRIOR FILING DATE: 2002-09-18  
PRIOR APPLICATION NUMBER: US 60/452,134  
PRIOR FILING DATE: 2003-03-04  
PRIOR APPLICATION NUMBER: US 60/411,537  
PRIOR FILING DATE: 2002-09-18  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 29  
LENGTH: 742  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: fusion protein  
US-10-418-836-29

Query Match 98.0%; Score 2714.5; DB 15; Length 742;  
Best Local Similarity 97.6%; Pred. No. 4,86-232;  
Matches 526; Conservative 2; Mismatches 6; Indels 5; Gaps 1;

QY 1 MSFRSLALSGVCTGLANVISKRATLDSMLNEATVARTAILNNIGADGAWSGADSGI 60  
DB 1 MSFRSLALSGVCTGLANVISKRATLDSMLNEATVARTAILNNIGADGAWSGADSGI 60  
QY 61 VVASPTDNDPYFYTWTRDSGLVLTVDLFRNGDTSLLSTIENYISAQAIVOGISNPSG 120  
DB 61 VVASPTDNDPYFYTWTRDSGLVLTVDLFRNGDTSLLSTIENYISAQAIVOGISNPSG 120  
QY 121 DISSGAGLGEPRKFNDEYATYSGWGRPORDGALRATAMIGFQWLLDNGYSTATDIYW 180  
DB 121 DISSGAGLGEPRKFNDEYATYSGWGRPORDGALRATAMIGFQWLLDNGYSTATDIYW 180  
QY 121 DISSGAGLGEPRKFNDEYATYSGWGRPORDGALRATAMIGFQWLLDNGYSTATDIYW 180  
DB 121 DISSGAGLGEPRKFNDEYATYSGWGRPORDGALRATAMIGFQWLLDNGYSTATDIYW 180  
QY 181 PLVRNDLSVYAQWNOGTGDLMEEVNGSSFITIAVOHRAIVEGSAFATVAGSSCWCDSQ 240  
DB 181 PLVRNDLSVYAQWNOGTGDLMEEVNGSSFITIAVOHRAIVEGSAFATVAGSSCWCDSQ 240  
QY 241 APEILCYLQSFMTGSFILANFDSRSRSGKDNATLLGSIHTEPDAACDDSTFOPCSPRALA 300  
DB 241 APEILCYLQSFMTGSFILANFDSRSRSGKDNATLLGSIHTEPDAACDDSTFOPCSPRALA 300  
QY 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNPFICTIAAEQLYDALYQMD 360  
DB 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNPFICTIAAEQLYDALYQMD 360  
QY 361 KQGSLEVTQVSLDFFKALYSDATGTYSSSSSTYSSIVDAVKTFADGFVSIYETHAASNG 420  
DB 361 KQGSLEVTQVSLDFFKALYSDATGTYSSSSSTYSSIVDAVKTFADGFVSIYETHAASNG 420  
QY 421 SMSEQYDKSDGBOLSRDLTWGYAALLTANRRNSVVPASWGETSASVPGTCAATSAIG 480  
DB 421 SMSEQYDKSDGBOLSRDLTWGYAALLTANRRNSVVPASWGETSASVPGTCAATSAIG 480  
QY 481 TVSSVTVTSWPSIVATGTTTATPTGSGSVTSTSKTTATASXTST-----TTRSGMSL 534  
DB 481 TVSSVTVTSWPSIVATGTTTATPTGSGSVTSTSKTTATASXTST-----TTRSGMSL 534

RESULT 20  
US-10-418-836-31  
Sequence 31, Application US/10418836  
Publication No. US20040018573A1  
GENERAL INFORMATION:  
APPLICANT: Power, Scott D.  
APPLICANT: Wang, Huang  
APPLICANT: Ward, Michael  
TITLE OF INVENTION: Production of Functional Antibodies in  
TITLE OF INVENTION: Filamentous Fungi  
FILE REFERENCE: GC741-2

CURRENT APPLICATION NUMBER: US/10/418,836  
CURRENT FILING DATE: 2003-04-17  
PRIOR APPLICATION NUMBER: US 60/373,889  
PRIOR FILING DATE: 2002-04-18  
PRIOR APPLICATION NUMBER: US 60/411,540  
PRIOR FILING DATE: 2002-09-18  
PRIOR APPLICATION NUMBER: US 60/452,134  
PRIOR FILING DATE: 2003-03-04  
PRIOR APPLICATION NUMBER: US 60/411,537  
PRIOR FILING DATE: 2002-09-18  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 31  
LENGTH: 743  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: fusion protein  
US-10-418-836-31

Query Match 97.9%; Score 2714; DB 15; Length 743;  
Best Local Similarity 100.0%; Pred. No. 5,36-232;  
Matches 522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFRSLALSGVCTGLANVISKRATLDSMLNEATVARTAILNNIGADGAWSGADSGI 60  
DB 1 MSFRSLALSGVCTGLANVISKRATLDSMLNEATVARTAILNNIGADGAWSGADSGI 60  
QY 61 VVASPTDNDPYFYTWTRDSGLVLTVDLFRNGDTSLLSTIENYISAQAIVOGISNPSG 120  
DB 61 VVASPTDNDPYFYTWTRDSGLVLTVDLFRNGDTSLLSTIENYISAQAIVOGISNPSG 120  
QY 121 DISSGAGLGEPRKFNDEYATYSGWGRPORDGALRATAMIGFQWLLDNGYSTATDIYW 180  
DB 121 DISSGAGLGEPRKFNDEYATYSGWGRPORDGALRATAMIGFQWLLDNGYSTATDIYW 180  
QY 121 DISSGAGLGEPRKFNDEYATYSGWGRPORDGALRATAMIGFQWLLDNGYSTATDIYW 180  
DB 121 DISSGAGLGEPRKFNDEYATYSGWGRPORDGALRATAMIGFQWLLDNGYSTATDIYW 180  
QY 181 PLVRNDLSVYAQWNOGTGDLMEEVNGSSFITIAVOHRAIVEGSAFATVAGSSCWCDSQ 240  
DB 181 PLVRNDLSVYAQWNOGTGDLMEEVNGSSFITIAVOHRAIVEGSAFATVAGSSCWCDSQ 240  
QY 241 APEILCYLQSFMTGSFILANFDSRSRSGKDNATLLGSIHTEPDAACDDSTFOPCSPRALA 300  
DB 241 APEILCYLQSFMTGSFILANFDSRSRSGKDNATLLGSIHTEPDAACDDSTFOPCSPRALA 300  
QY 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNPFICTIAAEQLYDALYQMD 360  
DB 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPEDTYNGNPFICTIAAEQLYDALYQMD 360  
QY 361 KQGSLEVTQVSLDFFKALYSDATGTYSSSSSTYSSIVDAVKTFADGFVSIYETHAASNG 420  
DB 361 KQGSLEVTQVSLDFFKALYSDATGTYSSSSSTYSSIVDAVKTFADGFVSIYETHAASNG 420  
QY 421 SMSEQYDKSDGBOLSRDLTWGYAALLTANRRNSVVPASWGETSASVPGTCAATSAIG 480  
DB 421 SMSEQYDKSDGBOLSRDLTWGYAALLTANRRNSVVPASWGETSASVPGTCAATSAIG 480  
QY 481 TVSSVTVTSWPSIVATGTTTATPTGSGSVTSTSKTTATASXTST-----TTRSGMSL 522  
DB 481 TVSSVTVTSWPSIVATGTTTATPTGSGSVTSTSKTTATASXTST-----TTRSGMSL 522

RESULT 21  
US-10-418-836-35  
Sequence 35, Application US/10418836  
Publication No. US20040018573A1  
GENERAL INFORMATION:  
APPLICANT: Power, Scott D.  
APPLICANT: Wang, Huang  
APPLICANT: Ward, Michael  
TITLE OF INVENTION: Production of Functional Antibodies in  
TITLE OF INVENTION: Filamentous Fungi  
FILE REFERENCE: GC741-2  
CURRENT APPLICATION NUMBER: US/10/418,836

; CURRENT FILING DATE: 2003-04-17  
; PRIOR APPLICATION NUMBER: US 60/373,889  
; PRIOR FILING DATE: 2002-04-18  
; PRIOR APPLICATION NUMBER: US 60/411,540  
; PRIOR FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: US 60/452,134  
; PRIOR FILING DATE: 2003-03-04  
; PRIOR APPLICATION NUMBER: US 60/411,537  
; PRIOR FILING DATE: 2002-09-18  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 35  
; LENGTH: 743  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: fusion protein  
US-10-418-836-35

Query Match 97.9%; Score 2714; DB 15; Length 743;  
Best Local Similarity 100.0%; Pred. No. 5,3e-232;  
Matches 522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFRLSLALSGVCTGLANVISKRATLDSWLSNEATVARTALINNGADGAWSGADSGI 60  
DB 1 MSFRLSLALSGVCTGLANVISKRATLDSWLSNEATVARTALINNGADGAWSGADSGI 60  
QY 61 VVASPSTDNPDYFYTWTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIVOGINSNPSG 120  
DB 61 VVASPSTDNPDYFYTWTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIVOGINSNPSG 120  
QY 121 DLSGAGLGEPRKENVDETAATYGSWGRPQRDPALRATAMIGFGQMLDNGYSTATDIW 180  
DB 121 DLSGAGLGEPRKENVDETAATYGSWGRPQRDPALRATAMIGFGQMLDNGYSTATDIW 180  
QY 121 DLSGAGLGEPRKENVDETAATYGSWGRPQRDPALRATAMIGFGQMLDNGYSTATDIW 180  
DB 121 DLSGAGLGEPRKENVDETAATYGSWGRPQRDPALRATAMIGFGQMLDNGYSTATDIW 180  
QY 181 PLVRNDLSYVAQVWNOTGYDLMEEVNGSSPFTIAVOHRALVEGSAFATAVSSCSWCDSQ 240  
DB 181 PLVRNDLSYVAQVWNOTGYDLMEEVNGSSPFTIAVOHRALVEGSAFATAVSSCSWCDSQ 240  
QY 241 APEILCYLQSFMTGSFLLANFDSRSRGKANTLLGSIHTFPPEACDSTFQPCSPRALA 300  
DB 241 APEILCYLQSFMTGSFLLANFDSRSRGKANTLLGSIHTFPPEACDSTFQPCSPRALA 300  
QY 301 NHKEVDSFRSIIYTLNDGLSDSEAVAVGRYPEDTYNGNPFCTLAABEQLYDALYQMD 360  
DB 301 NHKEVDSFRSIIYTLNDGLSDSEAVAVGRYPEDTYNGNPFCTLAABEQLYDALYQMD 360  
QY 361 KQGSLEVTDVSLDFPKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFVSIVETHAASNG 420  
DB 361 KQGSLEVTDVSLDFPKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFVSIVETHAASNG 420  
QY 421 SMSEQYDKSDGEQLSARDLTWSYALLLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480  
DB 421 SMSEQYDKSDGEQLSARDLTWSYALLLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480  
QY 481 TYSSTVYTSWPSIVATGCTTTATPTGSGSVTSTSKTTAAS 522  
DB 481 TYSSTVYTSWPSIVATGCTTTATPTGSGSVTSTSKTTAAS 522

RESULT 22  
US-10-418-836-39  
; Sequence 39, Application US/10418836  
; Publication No. US20040018573A1  
; GENERAL INFORMATION:  
; APPLICANT: Power, Scott D.  
; APPLICANT: Wang, Huaming  
; TITLE OF INVENTION: Production of Functional Antibodies in  
; FILE REFERENCE: GC741-2  
; CURRENT APPLICATION NUMBER: US/10/418,836  
; CURRENT FILING DATE: 2003-04-17

; PRIOR APPLICATION NUMBER: US 60/373,889  
; PRIOR FILING DATE: 2002-04-18  
; PRIOR APPLICATION NUMBER: US 60/411,540  
; PRIOR FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: US 60/452,134  
; PRIOR FILING DATE: 2003-03-04  
; PRIOR APPLICATION NUMBER: US 60/411,537  
; PRIOR FILING DATE: 2002-09-18  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 39  
; LENGTH: 975  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: fusion protein  
US-10-418-836-39

Query Match 97.9%; Score 2714; DB 15; Length 975;  
Best Local Similarity 100.0%; Pred. No. 8e-232;  
Matches 522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFRLSLALSGVCTGLANVISKRATLDSWLSNEATVARTALINNGADGAWSGADSGI 60  
DB 1 MSFRLSLALSGVCTGLANVISKRATLDSWLSNEATVARTALINNGADGAWSGADSGI 60  
QY 61 VVASPSTDNPDYFYTWTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIVOGINSNPSG 120  
DB 61 VVASPSTDNPDYFYTWTRDSGLVLTVDLFRNGDTSLSSTIENYISAQAIVOGINSNPSG 120  
QY 121 DLSGAGLGEPRKENVDETAATYGSWGRPQRDPALRATAMIGFGQMLDNGYSTATDIW 180  
DB 121 DLSGAGLGEPRKENVDETAATYGSWGRPQRDPALRATAMIGFGQMLDNGYSTATDIW 180  
QY 121 DLSGAGLGEPRKENVDETAATYGSWGRPQRDPALRATAMIGFGQMLDNGYSTATDIW 180  
DB 121 DLSGAGLGEPRKENVDETAATYGSWGRPQRDPALRATAMIGFGQMLDNGYSTATDIW 180  
QY 181 PLVRNDLSYVAQVWNOTGYDLMEEVNGSSPFTIAVOHRALVEGSAFATAVSSCSWCDSQ 240  
DB 181 PLVRNDLSYVAQVWNOTGYDLMEEVNGSSPFTIAVOHRALVEGSAFATAVSSCSWCDSQ 240  
QY 241 APEILCYLQSFMTGSFLLANFDSRSRGKANTLLGSIHTFPPEACDSTFQPCSPRALA 300  
DB 241 APEILCYLQSFMTGSFLLANFDSRSRGKANTLLGSIHTFPPEACDSTFQPCSPRALA 300  
QY 301 NHKEVDSFRSIIYTLNDGLSDSEAVAVGRYPEDTYNGNPFCTLAABEQLYDALYQMD 360  
DB 301 NHKEVDSFRSIIYTLNDGLSDSEAVAVGRYPEDTYNGNPFCTLAABEQLYDALYQMD 360  
QY 361 KQGSLEVTDVSLDFPKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFVSIVETHAASNG 420  
DB 361 KQGSLEVTDVSLDFPKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFVSIVETHAASNG 420  
QY 421 SMSEQYDKSDGEQLSARDLTWSYALLLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480  
DB 421 SMSEQYDKSDGEQLSARDLTWSYALLLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480  
QY 481 TYSSTVYTSWPSIVATGCTTTATPTGSGSVTSTSKTTAAS 522  
DB 481 TYSSTVYTSWPSIVATGCTTTATPTGSGSVTSTSKTTAAS 522

RESULT 23  
US-10-228-063-45  
; Sequence 45, Application US/10228063  
; Publication No. US20030135885A1  
; GENERAL INFORMATION:  
; APPLICANT: Lananan, Mike  
; TITLE OF INVENTION: Self-processing Plants and Plant Parts  
; FILE REFERENCE: 109846.317  
; CURRENT APPLICATION NUMBER: US/10/228,063  
; CURRENT FILING DATE: 2002-12-12  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 45  
; LENGTH: 1095





Db 13 AFOAALGDPPLHEKSHDIKR-SVDSYIQTETPIAOKNLLCNIGASGCRASGAAGVV 71  
QY VASPSPTDNDFY-----FYWTBDSGLVLTIVDLFRNG-DTSL 99  
Db 72 VASPSASSDLDLXETIAOCVSLMAALTYISIDWYITRDAALVTAKIIVDEFTINDVITLQ 131  
QY 100 STIENYISAQAIVOGISNPSGDLSSGAGLGEPEKFNVDETAYTGSWGRPORDEPALRATM 159  
Db 132 NTIQAAVAAOAKLQGVSNPSGSLNSGAGLGEPEKFNVDLQOFTGAKWRPORDEPALRATM 191  
QY 160 IGFQGLVLDNGITSTLTIDVPLVRNDLSVVAQYNNQGTGDLMEEVNNGSFFFTIAVOHRA 219  
Db 192 IGGKMLVSNRGYADTFKSIIMPIVKNDLAVTAAQYNNNGTGDLMEEVNSSFFFTIAASHRA 251  
QY 220 IVEGSAFATAVSGSSGMCDSQAPELLCYIOSFWTGS-FILANFDSRSGKDATLIGSH 278  
Db 252 LVEGSAFATAVSGSSGACDIAPIOLLCPQOSFWNSGYYIISFNVRSGDINSVLTSIH 311  
QY 279 TEPDEAACDSTFQPCSPALANHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNG 338  
Db 312 NFDPAAGCVNFTQPCSDRALANHKVVDMSR-FMGVNSRTAGKAAVGRVADVYNG 370  
QY 339 NFWFCTLAAAEQLYALYQMDKQSLLEVTVSLDFPKALYSDAATGTYSSSSTYSIV 398  
Db 371 NFWYLATTLAAEQLYDAVYVWKQSGITVSTSLAFKDLVPSVSTGTYSSSSTYTAII 430  
QY 399 DAVKTFADGFSIVETHAASNGSFOYKSDDEQLSARDLTWSYAAALTANRRNSVVP 458  
Db 431 NAWYATDADPVDIVAYTBSDGSLAOPDSDGAPLSAHLTWSYASFSAARRAGIYV 490  
QY 459 ASMGETSASSVPQCAATSAIGTYSSVTVSWPSIVATGTTT-ATPTGSGS-----VT 512  
Db 491 PBMGASANSLSPPSCASATYASGYATATATISFPANLTPASTVTPPTQICGADHEVLV 550  
QY 513 STSKTTATASKT 524  
Db 551 FMEKVTTSYGQT 562

RESULT 28  
US-10-281-673-3  
; Sequence 3, Application US/10281673  
; Publication No. US20030148452A1  
; GENERAL INFORMATION:  
; APPLICANT: Randy M. Berka  
; APPLICANT: Michael W. Rey  
; APPLICANT: Kimberly Brown  
; TITLE OF INVENTION: Promoters For Expressing Genes In A  
; FILE REFERENCE: 5611.200-US  
; CURRENT APPLICATION NUMBER: US/10/281,673  
; PRIOR FILING DATE: 2002-10-28  
; PRIOR APPLICATION NUMBER: US/09/534,407  
; PRIOR FILING DATE: 2000-03-22  
; PRIOR APPLICATION NUMBER: 09/274,449  
; PRIOR FILING DATE: 1999-03-22  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: FaSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 581  
; TYPE: PRT  
; ORGANISM: Pseudium  
; US-10-281-673-3

Query Match 51.4%; Score 1425.5; DB 14; Length 581;  
Best Local Similarity 52.2%; Pred. No. 1.3e-117;  
Matches 273; Conservative 81; Mismatches 146; Indels 23; Gaps 6;

QY 9 LSGVLVCTGL--ANVI---SKRATLDSKLSNEATVARTAIINNGAGGAWVSGADSGIVVA 63  
Db 6 LVLGVALWMOQGVVSPSKDNLERITQADISIKIGVIANIGAGGAKQAQAGAPAVVA 65  
QY 64 SPSTNDPBYFYWTBDSGLVLTIVDLFRNGDTSLSTIENYISAQAIVOGISNPSGDL 123

Db 66 SPSKEDPDWYWTBDSALTYKLVVERFIHGDKSLQRIKIDEVYSAQAKLQGTNNPSSGPE 125  
QY 124 SGAGLGEPEKFNVDETAYTGSWGRPORDEPALRATMIGCGWMLDNGYSTANDIYVPLV 183  
Db 126 SG-GLGEPEKFNVLNFTGSWGRPORDEPALRATLTLAEBWILSHGERSKALNKVMPV 184  
QY 184 RNDLSVVAQYNNQGTGDLMEEVNNGSFFFTIAVOHRAVLEGSAPATVAVSGSSGMCDSQAPE 243  
Db 185 EKDLATTFKFNRTGTDLMEEVNGSFFFTLSASHRLVGGALAKKLGSCEPCVTNNAR 244  
QY 244 ILCYIOSFWTGSFILANF--SSRSGKDATLIGSHITFDEPAACDSTFQPCSPALAN 301  
Db 245 VILCFQTFWGTGVDSININVKDGRKGLDVNSILSSIHTEFDPNKCTDSTFQPCSPALAN 304  
QY 302 HKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNFWFCTLAAAEQLYALYQMDK 361  
Db 305 HRAVDSFRSIYGVNNRQGAAGVRSSEVYDGNFWYATTLAAAEQLYAAVYQMDK 364  
QY 362 QSLLEVTVSLDFPKALYSDAATGTYSSSSTYSIVDAVKTFPADGFSIVETHAASNGS 421  
Db 365 LGAIVYDDVLSLFFKDIYVRKYSKTYAKTKTYKEIIKAAKTYADGFVAVVQTYTRKDS 424  
QY 422 MSEQYDKSDGEOLASARDLTWSYAAALTANRRNSVVPASMGETSASSVPQCAATSAIGT 481  
Db 425 LAEQFDKSTGAPKASVHLTWSYAFVATERRDGIISPWGESSANKVPACQADACDT 484  
QY 482 -----YSSVTVTSWPSIVATGTTT-----TPTGSG 509  
Db 485 TTFPSVKNQVSSDQKYVAVGVTLELNSWSPDGLALTPSSSG 527

RESULT 29  
US-10-213-990-21  
; Sequence 21, Application US/10213990  
; Publication No. US20030082595A1  
; GENERAL INFORMATION:  
; APPLICANT: Bussey, Bo  
; APPLICANT: Jiang, Howard  
; APPLICANT: Storms, Reg  
; APPLICANT: Roemer, Terry  
; TITLE OF INVENTION: NUCLEIC ACIDS OF ASPERGILLUS FUMIGATUS ENCODING INDUSTRIAL  
; FILE REFERENCE: 10182-019-999  
; CURRENT APPLICATION NUMBER: US/10/213,990  
; CURRENT FILING DATE: 2002-08-05  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FaSeq for Windows Version 4.0  
; SEQ ID NO 21  
; LENGTH: 704  
; TYPE: PRT  
; ORGANISM: Aspergillus  
; US-10-213-990-21

Query Match 50.8%; Score 1408.5; DB 14; Length 704;  
Best Local Similarity 52.0%; Pred. No. 5.7e-116;  
Matches 268; Conservative 85; Mismatches 139; Indels 23; Gaps 5;

QY 27 LDSWLSNEATVARTAIINNGAGGAWVSGADSGIVASPSPTNDPBYFYWTBDSGLVLT 86  
Db 118 LVSWLAQETSYALDGLVNVGPGAKATGASSGIIIASPSQSNPDYTYWTBDAALTVKY 177  
QY 87 LVDLF-RNGDTSLSTIENYISAQAIVOGISNPSGDLSSGAGLGEPEKFNVDETAYTGSW 145  
Db 178 LVGSFAADDPRIQRIIEBIVESQATLQTVSNPSGSLSSG-GLGEPEKLRVDSAFHGSWG 236  
QY 146 RPDORDEPALRATMIGCGWMLDNGYSTATDIPVPLVRNDLSVVAQYNNQGTGDLMEEV 205  
Db 237 RPDORDEPALRATLILSYALMIDNGYFTVESIWIYIIONDSLREFPNNSFFDLMEEV 286  
QY 206 NGSFFFTIAVOHRAVLEGSAPATVAVSGSSGMCDSQAPELLCYIOSFWTGSFILANFSSR 265  
Db 297 RGSFFFTIAVOHRAKGAALQRLGKTCNSCOSQAPVYLCFIQYTWGSSIIANLYSDR 356







Db 143 STISSWIKKEGISRPMANINP-----PGSATGFIASLSTAGPPYYATMRDAALTS 197  
QY 85 KTLVDLFR---NGDTSLLSTENYISAQAIVOGISNBSGDLSSGAGLGEPEKFNVDETAYT 141  
Db 198 NVIYEVNNTTILSGKNTILANTLKDVTFPSVKQSTSTVCN-----CLGEPEKFNVDAGYT 251  
QY 142 GSNCRPRDRPALRATMIGFG--QMLDNGYTSATITVWPLVRNDLSYVAQVWNOGYD 200  
Db 252 GAWGRPQNDGPABEATFTFLFADSYLQTQKADASYVTGTLKPAIFKDLDYVWVWNSNCFD 311  
QY 201 LMEEVNGSSFFITIAVQHRALVEGSAFATAVGSS--CSMCDSOAPILICYOSFWTGSFIL 258  
Db 312 LMEEVNGVHFTYTLWVWKGILLADPAKRGDSIRASTYSTASTANKISSEFWVSSNNW 371  
QY 259 ANFDS-----SSSGKADANTL---LGSHTFDEPAACDDSTFOPCSPRALANHKEVDS 308  
Db 372 IQVSQSVTVGVSKKGLDVSTLLANLGSV-----DDGFPTPGSEKILATVAVEDS 422  
QY 309 FRGITYTLNDLSDBEAVAVGRYPEDTY-----YNGNPFICTLAAEQLYDALYQMDKOG 363  
Db 423 FASLYPIPNKMLPSYLGNSISIRYPEDTYNGNSQNSWFLAVGTGYAELYYRAIKEWINGG 482  
QY 364 SLEVTDSLDFPKALYSDAATG--TYSSSSSTYSIYDAVTFADGFVSIYETHAASNGSM 422  
Db 483 GVTYVSSISLPEFKKFDSSATSCKYTVGTSDFNMLAONIALADRFSTVQLHANHNSGL 542  
QY 423 SEQYDKSDGEQLSARDLTWGSYALLTTAN 450  
Db 543 ABEFDRITGLSTGARDLTWSHSLITNS 570

RESULT 35  
US-10-369-493-22096  
; Sequence 22096, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 22096  
; LENGTH: 549  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-22096

Query Match 21.5%; Score 597; DB 15; Length 549;  
Best Local Similarity 30.4%; Pred. No. 5,5e-44;  
Matches 153; Conservative 89; Mismatches 202; Indels 60; Gaps 14;

QY 1 MSFRSLALSLGVLCTGLANYSKRATLDSWLSNEATVARTALINNIGADGAWTSGADSGI 60  
Db 56 VQURDAVLMWGTVAVD--SNAMDSLSALEWLQCKKXSIKIFENIGPSAVPS--ISPGV 113  
QY 61 VVASPTDNDPYFTYWTTRDSGLVLTVDLFRNGDTSLLSTIENYISAQAIVOGISNPSG 120  
Db 114 VINSPEQTHDPTYQWLRDSALITNSIVS---HSAGAIITLQYLVNSHILQPSNN--- 167  
QY 121 DLSGAG-----LGEPEKFNVDETAVTSGWRPQRDGPALRATMIGFGQWILNDG--- 170  
Db 168 TLGAGIGYTDVVALGDPKKNVNDTAFTEBDMGRPQNDGPALRSALILIKIIDYIKQSGTDL 227  
QY 171 -----YTSATIDVWPLVRNDLSYVAQVWNOGYDLMEEVNGSSFFITIAVQHRALVEGSA 225

Db 228 GAKTTPQSTRA-DIFDITVRMDLAFITIDHNSSGCFDLMEEVNGMHFFTLVQLSAVDKSL 286  
QY 226 FATAVGSSGSCWD---SOAPILICYL-----QSFMTGSFILANF-----DSRSKQDA 270  
Db 287 YFNASRRSSPFVELELQTRRIDISKFLVDPANGFINKY---NVIYGTPMIADTLRSGLDI 343  
QY 271 NTLGSIHTFDEPAACDDSTFOPCSPRALANHKEVDSRSIYTLNDGLSDSAVAVGRY 330  
Db 344 STLLANVTYHDAFSA--SHLFFDINDPAVNTLHLHLHMRISIPIINDSSKNATGIALGRY 402  
QY 331 PEDTY-----YNGNPFICTLAAEQLYDALYQMDKOGSLEVTDSLDFPKALYSDAATG 385  
Db 403 PEDVYGYGRGSGNPVLTCTHASTLYQILTRHISEQHDVVPANNDCSNAFWMSLIVS 462  
QY 386 TYSS-----SSSTYSIYDAVTFADGFVSIYETHAASNGSMSEQYDKSDGBO 433  
Db 463 NLTTLLNDSGYILLENTTAPFNQTIKIQGLADSLVKLKAHVGTGDELSEQKNKYTGFM 522  
QY 434 LSARDLTWGSYALLTTANRRNSVY 457  
Db 523 QGAQHLTWSYTSFMDAYQIRQEVLT 546

RESULT 36  
US-09-815-242-12713  
; Sequence 12713, Application US/09815242  
; Patent No. US2002061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12713  
; LENGTH: 2344  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12713

Query Match 5.9%; Score 164.5; DB 9; Length 2344;  
Best Local Similarity 20.6%; Pred. No. 0.00014;  
Matches 116; Conservative 103; Mismatches 234; Indels 111; Gaps 19;

QY 7 LALSGLVCTGLANYSKRATLDSWLSNEATVARTALINNIGADGAWVSG--ADSGIVAS 64  
Db 829 LVNSQSVSSSMGSGVSKTSLSDFINSSTSEKSESVSTSDSLKTSLSDSVSMST 888  
QY 65 PSTDNDPYFTYWTTRDSGLVLTVDLFRNGDTSLLSTIENYISAQAIVOGISNPSGDLSS 124



```

Db      889  GSLSKSGSLSTSDSASTSQSVSDSTSN-----1STSESLSESGOSTSEIS-1SN1SN 943
Qy      125  GAGLGEPEFENDETAIYCSMKRFPORDGALATAMIGCQWLDNGYST1DTIWPVLVR 184
Db      944  SVSASTSLSESGOSTS1SLSTS-----DCKSMST:- 972
Qy      185  NDLASVAYQWYNOTGYDLWEEVNGSSFFTIAYQHRALYEGSAFATAYVSGSCWCDQAPET 244
Db      973  -----ESLSDSTSTSDSVSGSLSVAGS---QSVSTSTSDSMSTSE 10099
Qy      245  LCVLOSFWTGSF-----TLANFDSRSRGKDMATLLGSIHTFPDEPA-ACDDSTPOPC 294
Db      1010  MIDSMSTSGSLAASDKMSKMSVSSSMSTSGSGSTSESLSDSI1STSDSPSKSLST1SGSG 10659
Qy      295  SPRLANHKEY---DSFRS1YTLTLDGLSDSEAVAVGYRPEBDYYNGNPFMLCTLAARQ 351
Db      1070  STSTSTSTSSSVRMSSESGOSTSGSMSTSQSDSTSLSTS--FSDSTSDSKS---ASTASSES 11244
Qy      352  LYDALWQMDKQGSU-----EYVDVSLDFPKALYSDAAATGYSSSSSTYSYDA 400
Db      1125  ISQSV-STSTSGSVSTSTSTSLSTSNSERTSTSMSTSDSTSLSTESDST-SDSTSTSDSISEA 11822
Qy      401  VKTPADGFSIVETHAASNG-----SMSEQYDKSDGEQLSADLTWVSYAALLTA 449
Db      1183  ISGSEBSTS1SLSEBNSTSDSKSASAFLESLSGSESTSESLSGS--TSDSTSLSDS 12440
Qy      450  NNRNRNVVPASWGETSASSVPGTCANISAICTYSSSVTYTSPSIYANGCTTATPTGSG 509
Db      1241  NSESGSTSTSLSNSTSGSA-----S1STSTSGSAGSTSVKS--ESVSTSLSTSTSTSLSDS 12944
Qy      510  SVTSTSKTTATASKTSTTRBGMS 533
Db      1295  TSLSTSLSDSTSGSKSNLSASMS 1318

RESULT 37
US-09-821-616-4
; Sequence 4, Application US/09821616
; Publication No. US20030027290A1
; GENERAL INFORMATION:
; APPLICANT: Nielsen, Bjarne R.
; APPLICANT: Nielsen, Ruby
; APPLICANT: Lehmebeck, Jan
; TITLE OF INVENTION: Thermostable Glucoamylase
; FILE REFERENCE: 5279.200-US
; CURRENT APPLICATION NUMBER: US/09/821,616
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/199,290
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 1557/97
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-12-30
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 0925/98
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/070,746
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/094,344
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/979,673
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/107,657
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: SeqID for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 35
; TYPE: prt
; ORGANISM: Talaromyces emersonii
US-09-821-616-4

```

[illegible]



Db 474 PA-----NGTSGTVEVEPT-----AGTITETIVSGSKTFTSPASGT-TS 515  
Qy 407 GFVSIYETHAASGMSSEQ---DKSDGRLSARDLTMSAALL--TANNRNSVVPAM 461  
Db 516 GIVEVEPEPA---GTTETIVSGSKAFSTFPANGTSGTVEVEPTAGTITETIVSGSV 572  
Qy 462 GTSASVPGTCATSAIGTYSSTVT-----SMPISVATGTTTATPFGSG 509  
Db 573 GYTSFTPAGTISGTYVEVEPTAGTITETIVSGSVGTSTFPASGTTSGTVEVEPT-AG 631  
Qy 510 SVTSTSKTATATASKISTTRSG 531  
Db 632 TTTETI-VSGSVGYTSTFPASG 652

RESULT 41  
US-10-185-990-10  
; Sequence 10, Application US/10185990  
; Publication No. US20030073109A1  
; GENERAL INFORMATION:  
; APPLICANT: Pan, Jae-Gu  
; TITLE OF INVENTION: JAE GU PAN ET AL  
; FILE REFERENCE: 02589.000100  
; CURRENT APPLICATION NUMBER: US/10/185,990  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 1621  
; TYPE: PRF  
; ORGANISM: BACILLUS SUBTILIS  
US-10-185-990-10

Query Match 5.6%; Score 155.5; DB 14; Length 1621;  
Best Local Similarity 23.6%; Pred. No. 0.00051;  
Matches 137; Conservative 66; Mismatches 245; Indels 133; Gaps 27;

Qy 32 SNEATVARTAILNINIGADG-----AMVSGADSGIVVASPTDNDP-----YFTWT 77  
Db 200 SNETAGNHSDDL IAGYSTGTAGYSTGTSGEDSLTAGYSTGTAGOGSNLTAGYSTGT 259  
Qy 78 --RDGSLVKTIVDLFRNGDTSLSTIENYISAQAIVOGISNPSGDLSSG-----A 126  
Db 260 AGSDSL IAGYSTGTSGEDSLTA---GYSTGTAGOGSNLTAGYSTGTAGVDSLLA 316  
Qy 127 GLGEPEFNDETAATYGSWGRPQRDGPALRATAMIGFGWLNDNGYSTATDIWPLVRND 186  
Db 317 GYSTGTSGEDSLTAGYSTGTQ-----TAQEGSN---LTAGYSTGT-----AGSD 359  
Qy 187 LSVTAQYMNQGTGYDLMEVNGSFFITIAV-QHRALVEGSAFATAVGSS-CSMCDQAPET 244  
Db 360 SSLLA-----GYSTGTSGEDSLTAGYSTGTAGOGSNLTAGYSTGTAGVDSLLAG 413  
Qy 245 LCVLQGFMTGSFLIANFDSRSRGKDNLT---LGSHTFPBAACDSTFQPC-SPRALA 300  
Db 414 YGSTGTSGEDSLTAGYSTGTAGOGSNLTAGYSTGT---AGADSSLIAGYSTGTSG 469  
Qy 301 NHKEVVDSPRSIYTLNDGL-----SDSEAVAVGRYPEDTYNGNPFPLCTIA 347  
Db 470 SSSSLTAGYSTGTAREGSTLTLAGYSTGTAGADSSLLA-GYSTGT-----S 516  
Qy 348 AAEQLYDALY---QMKQGSLEVTDVSLDFKALYSDAATGYSSSSSTYSIVDA--VK 402  
Db 517 GSSSSLTAGYSTGTAGOGSVLTSGYSTGTAGAAANLTTGYSTGTAGHESFLIAGYGS 576  
Qy 403 TPADEGVSIY-----ETHAASNGMSSEQYDKSDGELSARDL-----TMSYALLTA 449  
Db 577 TOTAGHKSILTAGYSTGTARDGSYLLIAGYSTGTAGSSSLIAGYSTGTATAYRMLTA 636  
Qy 450 -----NNRNSVVPAMSGETSASVPGTCATSAIGTYSSTVTSPISVATG-GTTTT 502  
Db 637 GYSTGTAREHSDLVLTGYGSTTAG-----SNSSLIAGYSTGTAGFKSILTAGYSTGT 691

Qy 503 ATP-----TSGSVTSTS-----XTATATASKISTTT 528  
Db 692 AGERSDLVAGYGS-ISTAGYSSSLIAGYSTGTAGHESFLT 731

RESULT 42  
US-10-185-990-11  
; Sequence 11, Application US/10185990  
; Publication No. US20030073109A1  
; GENERAL INFORMATION:  
; APPLICANT: Pan, Jae-Gu  
; TITLE OF INVENTION: JAE GU PAN ET AL  
; FILE REFERENCE: 02589.000100  
; CURRENT APPLICATION NUMBER: US/10/185,990  
; CURRENT FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 11  
; LENGTH: 1626  
; TYPE: PRF  
; ORGANISM: Bacillus subtilis  
US-10-185-990-11

Query Match 5.6%; Score 155.5; DB 14; Length 1626;  
Best Local Similarity 23.6%; Pred. No. 0.00051;  
Matches 137; Conservative 66; Mismatches 245; Indels 133; Gaps 27;

Qy 32 SNEATVARTAILNINIGADG-----AMVSGADSGIVVASPTDNDP-----YFTWT 77  
Db 200 SNETAGNHSDDL IAGYSTGTAGYSTGTSGEDSLTAGYSTGTAGOGSNLTAGYSTGT 259  
Qy 78 --RDGSLVKTIVDLFRNGDTSLSTIENYISAQAIVOGISNPSGDLSSG-----A 126  
Db 260 AGSDSL IAGYSTGTSGEDSLTA---GYSTGTAGOGSNLTAGYSTGTAGVDSLLA 316  
Qy 127 GLGEPEFNDETAATYGSWGRPQRDGPALRATAMIGFGWLNDNGYSTATDIWPLVRND 186  
Db 317 GYSTGTSGEDSLTAGYSTGTQ-----TAQEGSN---LTAGYSTGT-----AGSD 359  
Qy 187 LSVTAQYMNQGTGYDLMEVNGSFFITIAV-QHRALVEGSAFATAVGSS-CSMCDQAPET 244  
Db 360 SSLLA-----GYSTGTSGEDSLTAGYSTGTAGOGSNLTAGYSTGTAGVDSLLAG 413  
Qy 245 LCVLQGFMTGSFLIANFDSRSRGKDNLT---LGSHTFPBAACDSTFQPC-SPRALA 300  
Db 414 YGSTGTSGEDSLTAGYSTGTAGOGSNLTAGYSTGT---AGADSSLIAGYSTGTSG 469  
Qy 301 NHKEVVDSPRSIYTLNDGL-----SDSEAVAVGRYPEDTYNGNPFPLCTIA 347  
Db 470 SSSSLTAGYSTGTAREGSTLTLAGYSTGTAGADSSLLA-GYSTGT-----S 516  
Qy 348 AAEQLYDALY---QMKQGSLEVTDVSLDFKALYSDAATGYSSSSSTYSIVDA--VK 402  
Db 517 GSSSLTAGYSTGTAGOGSVLTSGYSTGTAGAAANLTTGYSTGTAGHESFLIAGYGS 576  
Qy 403 TPADEGVSIY-----ETHAASNGMSSEQYDKSDGELSARDL-----TMSYALLTA 449  
Db 577 TOTAGHKSILTAGYSTGTARDGSYLLIAGYSTGTAGSSSLIAGYSTGTATAYRMLTA 636  
Qy 450 -----NNRNSVVPAMSGETSASVPGTCATSAIGTYSSTVTSPISVATG-GTTTT 502  
Db 637 GYSTGTAREHSDLVLTGYGSTTAG-----SNSSLIAGYSTGTAGFKSILTAGYSTGT 691

RESULT 43  
US-09-815-242-10932  
; Sequence 10932, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:

```

; APPLICANT: Haeselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10932
; LENGTH: 1031
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
; US-09-815-242-10932

Query Match      5.6%; Score 154.5; DB 9; Length 1031;
Best Local Similarity 21.2%; Pred. No. 0.00031;
Matches 109; Conservative 69; Mismatches 206; Indels 129; Gaps 18;

QY 75 TWTTRDSGLVTKTVLDLFRNDTSLSTIENYISAQAVGSIINPSGDLSSGAG----- 127
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 469 TW---ESLISPTLYEQYKNTPIPRNTLATPV-ATTIVQGSIDPQOLVANTMDADAVP 524
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 128 -----LGEPRFNVDETAAYTGSWGRPQRDGPALATAMIGFGWLLDNGYTSTA 175
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 525 AAKIVNEVLANEIGQQKVTQVLTREYETS-----LVTNVNVPVTV 562
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 176 TDIWPLVVRNDLSVNAQYW-----NOTGYDIMEVUNG-----SFFPIIAVQHRALVE 222
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 563 ED---PAKKRDLSQLQTVASIGBANQYTVYSWQLFTAIGPKTIVPSYQOATQ--LLAE 617
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 223 GSAFATVAVSGSCWSCSOAPEILCYIOSFWTGSFILANPDSNRSGKANTLLGSIHTFDP 282
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 618 GQV-----SGDKTQGVQVQLTSNLOS--AMKVLVVKADITILERAFAENELASVHRLDE 668
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 283 EAACDSTFQPCSPRALANHKEVVDSEFRSIYTLNDG-----LSDEAV--AVGRY 330
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 669 SVYTTDSW-----QAMQDALIDT-----TTGEGSSKQLQLLAWSDBELLEPIIGGF 715
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 331 PERYTYNGNWPFLCTLAALAEQVLDALYQMKQGSLEVTDVSLDFPKALYSDAATGYSSS 390
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 716 K-----TPDADQKRINQLTQTKTALLLVEKSTETTSNTSSSTSSSTSS 762
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 391 SSTYSSIVDAVKTAFDGFVSIIVETHAASNGSMSEQVYKSDQQLSARDLTWSYAALLTAN 450
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 763 SSTSESTSTSTN-----ESSSTSESTSTSTNNESSSTSESSSTPSTSSSTSES 813
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 451 NRNSV-VPASWGETSASVPGTCATSAIGYSSVTVTSPSIATGCTTTATPTGSG 509
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 814 STPSTSESTSESTSTSTNNESSSTSESSSTPSTSTSESSSTPSTSSSTSES 873
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 510 SVTST-----SKTTATASKTSTTTTRSGMS 533
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

```

Db 874 STSSTSESTNNTNESWTPSTTSKTSSTSESSAS 906

RESULT 44
US-10-120-801-63
; Sequence 63, Application US/10120801
; Publication No. US20030203843A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shimkete, Richard
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Kekuda, Kamesh
; APPLICANT: Szytek, Kimberly
; APPLICANT: Mehriban, Fuad
; APPLICANT: Topper, James N.
; APPLICANT: Malvankar, Uriel
; APPLICANT: Wasserman, Scott
; APPLICANT: Balinger, Shlomit
; APPLICANT: Smithson, Glenda
; APPLICANT: Gunther, Erik
; APPLICANT: Komuves, Laszlo
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-340
; CURRENT APPLICATION NUMBER: US/10/120,801
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/285748
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286068
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/288334
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/291241
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/322284
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/285609
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 2117
; TYPE: PRT
; ORGANISM: Neuraepora crassa
; US-10-120-801-63

Query Match      5.5%; Score 152.5; DB 15; Length 2117;
Best Local Similarity 20.8%; Pred. No. 0.0014;
Matches 124; Conservative 83; Mismatches 251; Indels 137; Gaps 25;

QY 29 SWLSNATYARTAILNINIGDAGWVGADSGIVVASPSDNDPQYFTWTRD-----SGLYL 84
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 798 SVLTGSANSOPILALHSTSEGGWTV-----IPFCNDPQTCFTFNIDMGTSPGLTT 848
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 85 KTLVDFRN-GDTSLSSTIENYISAQAVGSIINPSGDLSSGAGL-----G 129
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 849 TEVTIRFTMKGSGDLW--IDKSKRPWGSVLGAQNPSSDLFEGWVIADKGSSEATLFFTPG 906
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 130 EPEFNVDETAAYTGSWGRPQRDGPALATAMIGFGWLLDNGYISTVTDI--VWPLVRNDL 187
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 907 AAPLNDPPIVSGAWTLNVND-----LTFGVHVVFIQTLRAIKVGPFLPDGSARF 957
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 188 SYVAQYWMQGYDL-----WEVNGSSFFPIAVQHRALVEGSAFATA-----VGS 232
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 958 KYLGCTYRDSANRLETTQAQPSDNDNGK-----COYAIITKKAFAAGQYIYEECVGR 1011
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 233 SCWSCSOAPEILCYL-----QSF--WTGSFILANPDSNRSGKANTLLGSIHTFDP 283
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1012 STPSPAKLVKDYDVCNFIYICGDKSGFCGCGVGSYMMWMTTGTGVFENGLTAPFRPASK 1071
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 284 AACDSTFQPC-----SPRALANHKEVVD-----SFRSL-----YT 314
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

```
Db      1072 AVYGDWEYAGCRTDNASPATRALNDRIYQGSSTNTIESCAQACAGFSFFGVEMGVCEYC 1131
QY      315 LNDGLSDSEAVAVNGRYEPT---YYNGNWFELCTLAAEOLYDALYQMDKQSLFETDV 371
Db      1132 GNN-LNPGSTVA---DEKTCNVCGDPTLGGGGRISVY-----KQKGTIVGNPS 1179
QY      372 LDFPKALYDADATGTYSSTYSIYDAVKTPAD-----GFVSIVETHAANGSMSEQ 425
Db      1180 TGVSSS--SGTASGTASATASASTSSSAIASGTPGNPOSIGQYSSLCGYSDAVASRLQ 1237
QY      426 YKSDGEOLASBDLT---WSY-----AALLTANRRNSVVPASWGETSA---SSVPG 471
Db      1238 GKNTGSNMVSLDDCATYCAGYKFGTEYSAECFCGNDLNGAAPVTDGRCNMLCNGNQO 1297
QY      472 TCATSAITGTYSVYVTSWPSIVATGTTTATPTGSGSVTSKTTATASKTST 526
Db      1298 ICGSGNGLSMYQLNPNGTSSSVTASGSATQASATASGTASSTALATTTSS 1352
```

## RESULT 45

```
US-10-228-063-53
; Sequence 53, Application US/10228063
; Publication No. US20030135865A1
; GENERAL INFORMATION:
; APPLICANT: Lanahan, Mike
; TITLE OF INVENTION: Self-processing Plants and Plant Parts
; FILE REFERENCE: 109846.317
; CURRENT APPLICATION NUMBER: US/10/228,063
; CURRENT FILING DATE: 2002-12-12
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 53
; LENGTH: 45
; TYPE: PRY
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic
US-10-228-063-53
```

```
Query Match      5.5%; Score 152; DB 14; Length 45;
Best Local Similarity 88.9%; Pred. No. 4.5e-06;
Matches 32; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY      495 ATGGTTTATPTGSGSVTSKTTATASKTSTTRS 530
Db      1 ATGGTTTATPTGSGSVTSKTTATASKTSTTSS 36
```

Search completed: June 28, 2004, 07:50:52  
Job time : 54 secs

**THIS PAGE LEFT BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 07:42:07 ; Search time 23 Seconds  
(without alignments)  
1198.620 Million cell updates/sec

Title: US-10-038-723-2

Perfect score: 2771

Sequence: 1 MSFRSLALSGLVCTGLANV.....SKTTATSKYSTTSGMSL 534

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 100 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2771	100.0	534	3	US-09-198-672-2 Sequence 2, Appl1
2	2771	100.0	534	3	US-09-199-290-9 Sequence 9, Appl1
3	2771	100.0	534	4	US-09-632-392-2 Sequence 2, Appl1
4	2771	100.0	534	4	US-09-455-679-1 Sequence 1, Appl1
5	2771	100.0	534	4	US-09-351-814-2 Sequence 2, Appl1
6	2771	100.0	534	4	US-09-821-616-9 Sequence 9, Appl1
7	2628	99.0	640	4	US-09-351-814-13 Sequence 13, Appl1
8	2628	94.8	616	4	US-09-236-063-1 Sequence 1, Appl1
9	1689	61.0	618	3	US-09-199-290-34 Sequence 34, Appl1
10	1689	61.0	618	4	US-09-821-616-34 Sequence 34, Appl1
11	1674.5	60.4	531	3	US-09-199-290-7 Sequence 7, Appl1
12	1674.5	60.4	531	4	US-09-821-616-7 Sequence 7, Appl1
13	1615.5	58.3	630	4	US-09-704-449-2 Sequence 2, Appl1
14	1577	56.9	626	2	US-08-596-300A-7 Sequence 7, Appl1
15	1577	56.9	626	2	US-08-596-300A-14 Sequence 14, Appl1
16	1425.5	51.4	581	4	US-09-534-407-3 Sequence 3, Appl1
17	1425.5	51.4	581	4	US-09-999-201B-4 Sequence 4, Appl1
18	1321	47.7	616	1	US-08-385-370-2 Sequence 2, Appl1
19	1321	47.7	616	1	US-08-385-370-4 Sequence 4, Appl1
20	589	21.3	806	1	US-08-270-076A-11 Sequence 11, Appl1
21	160	5.8	35	3	US-09-199-290-4 Sequence 4, Appl1
22	160	5.8	35	4	US-09-821-616-4 Sequence 4, Appl1
23	156.5	5.6	2137	4	US-09-114-001C-4463 Sequence 4463, Ap
24	142	5.1	536	4	US-09-463-712C-10 Sequence 10, Appl1
25	136	4.9	1026	2	US-08-614-377A-7 Sequence 7, Appl1
26	136	4.9	1026	3	US-09-142-648B-7 Sequence 7, Appl1
27	133	4.8	1529	2	US-08-728-470-10 Sequence 10, Appl1

28	133	4.8	1529	3	US-08-719-641-10 Sequence 10, Appl1
29	132	4.8	1426	3	US-09-136-574A-43 Sequence 43, Appl1
30	130.5	4.7	1537	1	US-08-325-267A-2 Sequence 2, Appl1
31	130	4.7	1600	2	US-08-617-697-10 Sequence 10, Appl1
32	129	4.7	1026	1	US-08-194-290-7 Sequence 7, Appl1
33	128.5	4.6	673	4	US-09-107-532A-5134 Sequence 5134, Ap
34	128	4.6	1721	3	US-08-700-651-5 Sequence 5, Appl1
35	128	4.6	1721	3	US-08-928-361B-6 Sequence 6, Appl1
36	128	4.6	1721	3	US-09-588-995A-6 Sequence 6, Appl1
37	127.5	4.6	751	6	5457037-5 Patent No. 5457037
38	127.5	4.6	776	3	US-09-346-237-7 Sequence 7, Appl1
39	127	4.6	630	2	US-08-797-366-3 Sequence 3, Appl1
40	127	4.6	630	2	US-08-956-268-3 Sequence 3, Appl1
41	127	4.6	906	4	US-09-717-364A-15 Sequence 15, Appl1
42	126.5	4.6	776	3	US-09-346-237-4 Sequence 4, Appl1
43	126	4.5	288	4	US-09-216-393B-341 Sequence 341, App
44	126	4.5	288	4	US-09-216-393B-344 Sequence 344, App
45	124.5	4.5	894	3	US-08-362-525-22 Sequence 22, Appl1
46	124.5	4.5	894	3	US-08-971-692-15 Sequence 15, Appl1
47	124.5	4.5	3892	4	US-09-328-352-5503 Sequence 5503, Ap
48	123	4.4	562	6	5258502-2 Patent No. 5258502
49	123	4.4	725	2	US-08-816-105A-1 Sequence 1, Appl1
50	122.5	4.4	750	6	5457037-3 Patent No. 5457037
51	122	4.4	862	1	US-08-325-267A-4 Sequence 4, Appl1
52	121.5	4.4	2411	4	US-09-268-347-36 Sequence 36, Appl1
53	121	4.4	629	4	US-09-241-581B-6 Sequence 6, Appl1
54	121	4.4	629	4	US-08-265-428-6 Sequence 6, Appl1
55	121	4.4	629	5	PCT-US95-07721-6 Sequence 6, Appl1
56	120.5	4.3	823	4	US-09-107-532A-5667 Sequence 5667, Ap
57	120	4.3	1338	2	US-08-719-641-9 Sequence 9, Appl1
58	120	4.3	1338	3	US-08-719-641-9 Sequence 9, Appl1
59	120	4.3	1599	2	US-08-617-697-9 Sequence 9, Appl1
60	119.5	4.3	669	4	US-09-107-532A-6532 Sequence 6532, Ap
61	119	4.3	693	1	US-08-056-200-111 Sequence 111, App
62	119	4.3	693	2	US-08-800-644-111 Sequence 111, App
63	119	4.3	2736	2	US-09-252-991A-30227 Sequence 30227, A
64	118	4.3	52	6	5198345-14 Patent No. 5198345
65	117.5	4.2	1912	1	US-08-409-995-4 Sequence 4, Appl1
66	117.5	4.2	1912	3	US-08-685-467-4 Sequence 4, Appl1
67	117.5	4.2	2039	4	US-09-077-098A-7 Sequence 7, Appl1
68	117.5	4.2	2353	3	US-09-377-155-33 Sequence 33, Appl1
69	117.5	4.2	2353	3	US-08-913-942-4 Sequence 4, Appl1
70	117.5	4.2	2353	4	US-09-669-974-33 Sequence 33, Appl1
71	117.5	4.2	2353	4	US-09-797-862-33 Sequence 33, Appl1
72	117.5	4.2	2354	4	US-09-268-347-47 Sequence 47, Appl1
73	117	4.2	1004	4	US-09-206-942-57 Sequence 57, Appl1
74	117	4.2	1010	4	US-09-206-942-55 Sequence 55, Appl1
75	116	4.2	413	1	US-08-295-676A-2 Sequence 2, Appl1
76	116	4.2	413	1	US-08-750-077-6 Sequence 6, Appl1
77	116	4.2	413	2	US-08-948-591-2 Sequence 2, Appl1
78	116	4.2	1388	4	US-09-543-681A-5641 Sequence 5641, Ap
79	115.5	4.2	270	4	US-09-134-000C-3553 Sequence 3553, Ap
80	115	4.1	657	1	US-09-543-681A-7109 Sequence 7109, Ap
81	114.5	4.1	617	1	US-08-361-920-29 Sequence 29, Appl1
82	114.5	4.1	617	1	US-08-479-939-29 Sequence 29, Appl1
83	114.5	4.1	617	1	US-08-483-432-29 Sequence 29, Appl1
84	114.5	4.1	1751	3	US-09-136-574A-44 Sequence 44, Appl1
85	114	4.1	24	6	5198345-26 Patent No. 5198345
86	113.5	4.1	550	4	US-09-198-452A-613 Sequence 613, Appl1
87	113.5	4.1	573	6	5198345-19 Patent No. 5198345
88	113	4.1	28	6	5198345-19 Patent No. 5198345
89	113	4.1	553	1	US-08-651-542-2 Sequence 2, Appl1
90	113	4.1	553	3	US-09-066-544-2 Sequence 2, Appl1
91	113	4.1	553	3	US-08-951-086-2 Sequence 2, Appl1
92	113	4.1	553	3	US-09-430-669-2 Sequence 2, Appl1
93	113	4.1	715	4	US-09-620-412C-321 Sequence 321, App
94	113	4.1	715	4	US-09-598-419-321 Sequence 321, App
95	113	4.1	1060	4	US-08-811-393-2 Sequence 2, Appl1
96	112.5	4.1	915	4	US-09-206-942-35 Sequence 35, Appl1
97	112.5	4.1	1222	4	US-09-206-942-37 Sequence 37, Appl1
98	112.5	4.1	1228	4	US-09-206-942-34 Sequence 34, Appl1
99	112.5	4.1	1837	3	US-08-928-361B-5 Sequence 5, Appl1
100	112.5	4.1	1837	4	US-09-588-995A-5 Sequence 5, Appl1

## ALIGNMENTS

## RESULT 1

```
US-09-198-672-2
; Sequence 2, Application US/09198672
; Patent No. 6129788
; GENERAL INFORMATION:
; APPLICANT: Liaw, Gin
; APPLICANT: Pedersen, Sven
; APPLICANT: Hendiksen, Sven
; TITLE OF INVENTION: A Method of Producing Saccharide
; FILE REFERENCE: 5318-200-US
; CURRENT APPLICATION NUMBER: US/09/198, 672
; CURRENT FILING DATE: 1998-11-23
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Aspergillus Niger
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-09-198-672-2
```

Query Match 100.0%; Score 2771; DB 3; Length 534;

Best Local Similarity 100.0%; Pred. No. 1,1e-231; Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MSFRSLALSGVCTGLANVYSKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
DB 1 MSFRSLALSGVCTGLANVYSKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
QY 61 VVASPTDNDPYFYTWTRDSGLVLTIVLFRNGDTSLSTIENYISAQIIVOGISNPSG 120
DB 61 VVASPTDNDPYFYTWTRDSGLVLTIVLFRNGDTSLSTIENYISAQIIVOGISNPSG 120
QY 121 DISSGAGLGEPRKNVDETATYTGSGWRPQRDGPALRAITAMIGFGOMLLDNGYSTATDIW 180
DB 121 DISSGAGLGEPRKNVDETATYTGSGWRPQRDGPALRAITAMIGFGOMLLDNGYSTATDIW 180
QY 121 DISSGAGLGEPRKNVDETATYTGSGWRPQRDGPALRAITAMIGFGOMLLDNGYSTATDIW 180
DB 121 DISSGAGLGEPRKNVDETATYTGSGWRPQRDGPALRAITAMIGFGOMLLDNGYSTATDIW 180
QY 181 PLVRNDLSYVAQYWNQGYDLMEEVNSSFETIAVOHRALVEGSAPATAVSSCSWCDSQ 240
DB 181 PLVRNDLSYVAQYWNQGYDLMEEVNSSFETIAVOHRALVEGSAPATAVSSCSWCDSQ 240
QY 241 APEILCYLOSFWTGSFTLANFDSRSGKDNATLLGSIHTFDPBAACDSTFQPCSPRALA 300
DB 241 APEILCYLOSFWTGSFTLANFDSRSGKDNATLLGSIHTFDPBAACDSTFQPCSPRALA 300
QY 301 NHKEVVDSPRSITTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTIAAEQLYDALYQMD 360
DB 301 NHKEVVDSPRSITTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTIAAEQLYDALYQMD 360
QY 361 KQGSLEVTDVSLDFFKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFVSIIVETHAASNG 420
DB 361 KQGSLEVTDVSLDFFKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFVSIIVETHAASNG 420
QY 421 SMSEQYDKSDGEOLASRDLTWSYAALLTANNRNSVVPASWGETSASVPGCTCAATSAG 480
DB 421 SMSEQYDKSDGEOLASRDLTWSYAALLTANNRNSVVPASWGETSASVPGCTCAATSAG 480
QY 481 TYSSTVVTSMPSIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTTSSGMSL 534
DB 481 TYSSTVVTSMPSIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTTSSGMSL 534
```

## RESULT 2

```
US-09-199-290-9
; Sequence 9, Application US/09199290
```

; Patent No. 6255084

; GENERAL INFORMATION:

; APPLICANT: Nielsen, Bjørne R.

; APPLICANT: Nielsen, Ruby

; APPLICANT: Lehmbeck, Jan

; TITLE OF INVENTION: Thermostable Glucoamylase

; FILE REFERENCE: 5279,200-US

; CURRENT APPLICATION NUMBER: US/09/199, 290

; CURRENT FILING DATE: 1998-11-24

; EARLIER APPLICATION NUMBER: 1557/97

; EARLIER FILING DATE: 1997-12-30

; EARLIER APPLICATION NUMBER: 0925/98

; EARLIER FILING DATE: 1998-07-10

; EARLIER APPLICATION NUMBER: 60/070, 746

; EARLIER FILING DATE: 1998-01-08

; EARLIER APPLICATION NUMBER: 60/094, 344

; EARLIER FILING DATE: 1998-07-28

; EARLIER APPLICATION NUMBER: 08/979, 673

; EARLIER FILING DATE: 1997-11-26

; EARLIER APPLICATION NUMBER: 09/107, 657

; EARLIER FILING DATE: 1998-06-30

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 9

; LENGTH: 534

; TYPE: PRT

; ORGANISM: Aspergillus niger

; FEATURE:

; NAME/KEY: SIGNAL

; LOCATION: (1)...(24)

US-09-199-290-9

Query Match 100.0%; Score 2771; DB 3; Length 534;

Best Local Similarity 100.0%; Pred. No. 1,1e-231; Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MSFRSLALSGVCTGLANVYSKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
DB 1 MSFRSLALSGVCTGLANVYSKRATLDSWLSNEATVARTAIINNIAGDAMVSGADSGI 60
QY 61 VVASPTDNDPYFYTWTRDSGLVLTIVLFRNGDTSLSTIENYISAQIIVOGISNPSG 120
DB 61 VVASPTDNDPYFYTWTRDSGLVLTIVLFRNGDTSLSTIENYISAQIIVOGISNPSG 120
QY 121 DISSGAGLGEPRKNVDETATYTGSGWRPQRDGPALRAITAMIGFGOMLLDNGYSTATDIW 180
DB 121 DISSGAGLGEPRKNVDETATYTGSGWRPQRDGPALRAITAMIGFGOMLLDNGYSTATDIW 180
QY 121 DISSGAGLGEPRKNVDETATYTGSGWRPQRDGPALRAITAMIGFGOMLLDNGYSTATDIW 180
DB 121 DISSGAGLGEPRKNVDETATYTGSGWRPQRDGPALRAITAMIGFGOMLLDNGYSTATDIW 180
QY 181 PLVRNDLSYVAQYWNQGYDLMEEVNSSFETIAVOHRALVEGSAPATAVSSCSWCDSQ 240
DB 181 PLVRNDLSYVAQYWNQGYDLMEEVNSSFETIAVOHRALVEGSAPATAVSSCSWCDSQ 240
QY 241 APEILCYLOSFWTGSFTLANFDSRSGKDNATLLGSIHTFDPBAACDSTFQPCSPRALA 300
DB 241 APEILCYLOSFWTGSFTLANFDSRSGKDNATLLGSIHTFDPBAACDSTFQPCSPRALA 300
QY 301 NHKEVVDSPRSITTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTIAAEQLYDALYQMD 360
DB 301 NHKEVVDSPRSITTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTIAAEQLYDALYQMD 360
QY 361 KQGSLEVTDVSLDFFKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFVSIIVETHAASNG 420
DB 361 KQGSLEVTDVSLDFFKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFVSIIVETHAASNG 420
QY 421 SMSEQYDKSDGEOLASRDLTWSYAALLTANNRNSVVPASWGETSASVPGCTCAATSAG 480
DB 421 SMSEQYDKSDGEOLASRDLTWSYAALLTANNRNSVVPASWGETSASVPGCTCAATSAG 480
QY 481 TYSSTVVTSMPSIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTTSSGMSL 534
DB 481 TYSSTVVTSMPSIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTTSSGMSL 534
```



```
RESULT 3
US-09-632-392-2
; Sequence 2, Application US/09632392
; Patent No. 6303346
; GENERAL INFORMATION:
; APPLICANT: L'aw, Gln
; APPLICANT: Pedersen, Sven
; APPLICANT: Hendriksen, Sven
; TITLE OF INVENTION: A Method of Producing Saccharide
; FILE REFERENCE: 5318.200-US
; CURRENT APPLICATION NUMBER: US/09/632.392
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 09/198.672
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FaastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Aspergillus Niger
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-09-632-392-2

Query Match      100.0%; Score 2771; DB 4; Length 534;
Best Local Similarity 100.0%; Pred. No. 1.1e-231;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFRSLALSLGLVCTGLANVYSKRATLDSWLSNEATVARTAIINNIGADGAWVSGADSGI 60
DB 1 MSFRSLALSLGLVCTGLANVYSKRATLDSWLSNEATVARTAIINNIGADGAWVSGADSGI 60
QY 61 VVASPSTNDPDYFYTYTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIVOGINSNPG 120
DB 61 VVASPSTNDPDYFYTYTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIVOGINSNPG 120
QY 121 DISSGAGLGEPRKNDVETAYTSGWRPQRDPALRATAMIGQWLLDNGYSTATDIYW 180
DB 121 DISSGAGLGEPRKNDVETAYTSGWRPQRDPALRATAMIGQWLLDNGYSTATDIYW 180
QY 121 DISSGAGLGEPRKNDVETAYTSGWRPQRDPALRATAMIGQWLLDNGYSTATDIYW 180
DB 121 DISSGAGLGEPRKNDVETAYTSGWRPQRDPALRATAMIGQWLLDNGYSTATDIYW 180
QY 181 PLVRNDLSVVAQYMNQGYDLMEEVNNGSFFTLAVQHRALVEGSAPATAVSSCSWCDSD 240
DB 181 PLVRNDLSVVAQYMNQGYDLMEEVNNGSFFTLAVQHRALVEGSAPATAVSSCSWCDSD 240
QY 241 APEILCYLOSFWTGSFILANFDSRSRSGKDANTLLGSIHTFDEPAACDDSTFOPCSPRALA 300
DB 241 APEILCYLOSFWTGSFILANFDSRSRSGKDANTLLGSIHTFDEPAACDDSTFOPCSPRALA 300
QY 301 NHKEVVDTSRISYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMD 360
DB 301 NHKEVVDTSRISYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMD 360
QY 301 NHKEVVDTSRISYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMD 360
DB 301 NHKEVVDTSRISYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMD 360
QY 361 KQGSLEVTDSVLDPEFALYSDAATGTYSSTSSSTYSIVDAVKTFAAGFVSIVETHAASNG 420
DB 361 KQGSLEVTDSVLDPEFALYSDAATGTYSSTSSSTYSIVDAVKTFAAGFVSIVETHAASNG 420
QY 421 SMSEQYDKSDGEOQLSARDLTWSYAALLTANNRNSVVPASWGETSASSVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGEOQLSARDLTWSYAALLTANNRNSVVPASWGETSASSVPGTCAATSAIG 480
QY 481 TVSSVTVTWSPISIVATGCTTTTATPTGSGSVSTSKTTATASKTSTTTRSGMSL 534
DB 481 TVSSVTVTWSPISIVATGCTTTTATPTGSGSVSTSKTTATASKTSTTTRSGMSL 534

RESULT 4
US-09-455-679-1
; Sequence 1, Application US/09455679
; Patent No. 6329186
; GENERAL INFORMATION:
; APPLICANT: Nielsen, Bjarne R.
```

```
; APPLICANT: Svendsen, Allan
; APPLICANT: Bojsen, Kirsten
; APPLICANT: Vind, Jesper
; APPLICANT: Pedersen, Henrik
; TITLE OF INVENTION: Glucoamylases With N-Terminal Extensions
; FILE REFERENCE: 5691.200-US
; CURRENT APPLICATION NUMBER: US/09/455.679
; PRIOR FILING DATE: 1999-12-07
; EARLIER APPLICATION NUMBER: PA 1998 01616
; EARLIER FILING DATE: 1998-12-07
; EARLIER APPLICATION NUMBER: PA 1999 00409
; EARLIER FILING DATE: 1999-03-24
; EARLIER APPLICATION NUMBER: 60/111,674
; EARLIER FILING DATE: 1998-12-10
; EARLIER APPLICATION NUMBER: 60/126,740
; EARLIER FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Aspergillus niger
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-09-455-679-1

Query Match      100.0%; Score 2771; DB 4; Length 534;
Best Local Similarity 100.0%; Pred. No. 1.1e-231;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MSFRSLALSLGLVCTGLANVYSKRATLDSWLSNEATVARTAIINNIGADGAWVSGADSGI 60
DB 1 MSFRSLALSLGLVCTGLANVYSKRATLDSWLSNEATVARTAIINNIGADGAWVSGADSGI 60
QY 61 VVASPSTNDPDYFYTYTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIVOGINSNPG 120
DB 61 VVASPSTNDPDYFYTYTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIVOGINSNPG 120
QY 121 DISSGAGLGEPRKNDVETAYTSGWRPQRDPALRATAMIGQWLLDNGYSTATDIYW 180
DB 121 DISSGAGLGEPRKNDVETAYTSGWRPQRDPALRATAMIGQWLLDNGYSTATDIYW 180
QY 121 DISSGAGLGEPRKNDVETAYTSGWRPQRDPALRATAMIGQWLLDNGYSTATDIYW 180
DB 121 DISSGAGLGEPRKNDVETAYTSGWRPQRDPALRATAMIGQWLLDNGYSTATDIYW 180
QY 181 PLVRNDLSVVAQYMNQGYDLMEEVNNGSFFTLAVQHRALVEGSAPATAVSSCSWCDSD 240
DB 181 PLVRNDLSVVAQYMNQGYDLMEEVNNGSFFTLAVQHRALVEGSAPATAVSSCSWCDSD 240
QY 241 APEILCYLOSFWTGSFILANFDSRSRSGKDANTLLGSIHTFDEPAACDDSTFOPCSPRALA 300
DB 241 APEILCYLOSFWTGSFILANFDSRSRSGKDANTLLGSIHTFDEPAACDDSTFOPCSPRALA 300
QY 301 NHKEVVDTSRISYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMD 360
DB 301 NHKEVVDTSRISYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMD 360
QY 301 NHKEVVDTSRISYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMD 360
DB 301 NHKEVVDTSRISYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMD 360
QY 361 KQGSLEVTDSVLDPEFALYSDAATGTYSSTSSSTYSIVDAVKTFAAGFVSIVETHAASNG 420
DB 361 KQGSLEVTDSVLDPEFALYSDAATGTYSSTSSSTYSIVDAVKTFAAGFVSIVETHAASNG 420
QY 421 SMSEQYDKSDGEOQLSARDLTWSYAALLTANNRNSVVPASWGETSASSVPGTCAATSAIG 480
DB 421 SMSEQYDKSDGEOQLSARDLTWSYAALLTANNRNSVVPASWGETSASSVPGTCAATSAIG 480
QY 481 TVSSVTVTWSPISIVATGCTTTTATPTGSGSVSTSKTTATASKTSTTTRSGMSL 534
DB 481 TVSSVTVTWSPISIVATGCTTTTATPTGSGSVSTSKTTATASKTSTTTRSGMSL 534

RESULT 5
US-09-351-814-2
; Sequence 2, Application US/09351814
; Patent No. 6352851
; GENERAL INFORMATION:
```

```

; APPLICANT: Nielsen, Bjarne Roenfeldt
; APPLICANT: Svendsen, Allan
; APPLICANT: Pedersen, Henrik
; APPLICANT: Vind, Jesper
; APPLICANT: Hendriksen, Hanne Vang
; APPLICANT: Frandsen, Torben Peter
; TITLE OF INVENTION: Glucoamylase Variants
; FILE REFERENCE: 5636.200-US
; CURRENT APPLICATION NUMBER: US/09/351,814
; EARLIER FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: PA 1998 00937
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: PA 1998 01667
; EARLIER FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/093,528
; EARLIER FILING DATE: 1998-07-21
; EARLIER APPLICATION NUMBER: 60/115,545
; EARLIER FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: FaetSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 534
; TYPE: PR1
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1) ... (24)
US-09-351-814-2
```

Query Match 100.0%; Score 2771; DB 4; Length 534;

Best Local Similarity 100.0%; Pred. No. 1.1e-231; Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 MSFRSLALSGVCTGLANVSKRATLDSMLSEATVARTALINNGADGAWSGADSGI 60
Db 1 MSFRSLALSGVCTGLANVSKRATLDSMLSEATVARTALINNGADGAWSGADSGI 60
Qy 61 VVASPSTNDPDYFTYTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIVOGINSNPSG 120
Db 61 VVASPSTNDPDYFTYTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIVOGINSNPSG 120
Qy 121 DLSGAGLGEPRKFNVDATATGSGWRPQRDPALRATAMIGFGQWLLDNGYSTATDIW 180
Db 121 DLSGAGLGEPRKFNVDATATGSGWRPQRDPALRATAMIGFGQWLLDNGYSTATDIW 180
Qy 121 DLSGAGLGEPRKFNVDATATGSGWRPQRDPALRATAMIGFGQWLLDNGYSTATDIW 180
Db 121 DLSGAGLGEPRKFNVDATATGSGWRPQRDPALRATAMIGFGQWLLDNGYSTATDIW 180
Qy 181 PLVRNDLSVYAOYWNQGYDLMEEVNSSFPTIAVQHRALVEGSARFATAVGSSCWCDSQ 240
Db 181 PLVRNDLSVYAOYWNQGYDLMEEVNSSFPTIAVQHRALVEGSARFATAVGSSCWCDSQ 240
Qy 241 APEILCYLQSFMTGSFTLANFDSRSKGDANTLLGSIHTFDPACDSTFOPCSPRALA 300
Db 241 APEILCYLQSFMTGSFTLANFDSRSKGDANTLLGSIHTFDPACDSTFOPCSPRALA 300
Qy 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPFCTTAAAEQLDALYQMD 360
Db 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPFCTTAAAEQLDALYQMD 360
Qy 361 KQGSLEVTVDVSLDFFKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFPISIVETHAASNG 420
Db 361 KQGSLEVTVDVSLDFFKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFPISIVETHAASNG 420
Qy 421 SMEQYDKSDGEOLSARDLTWSYALLTANRRNSVVPASWGETSASVPGTCAATSAIG 480
Db 421 SMEQYDKSDGEOLSARDLTWSYALLTANRRNSVVPASWGETSASVPGTCAATSAIG 480
Qy 481 TYSSVTVTSPSPSVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTRSGMSL 534
Db 481 TYSSVTVTSPSPSVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTRSGMSL 534
```

RESULT 6  
US-09-821-616-9  
Sequence 9, Application US/09821616

```

; Patent No. 6620924
; GENERAL INFORMATION:
; APPLICANT: Nielsen, Bjarne R.
; APPLICANT: Nielsen, Rudy
; APPLICANT: Lehmbeck, Jan
; TITLE OF INVENTION: Thermostable Glucoamylase
; FILE REFERENCE: 5279.200-US
; CURRENT APPLICATION NUMBER: US/09/821,616
; PRIOR APPLICATION NUMBER: 2001-03-29
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1557/97
; PRIOR FILING DATE: 1997-12-30
; PRIOR APPLICATION NUMBER: 0925/98
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: 60/070,746
; PRIOR FILING DATE: 1998-01-08
; PRIOR APPLICATION NUMBER: 60/094,344
; PRIOR FILING DATE: 1998-07-28
; PRIOR APPLICATION NUMBER: 08/979,673
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 09/107,657
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FaetSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 534
; TYPE: PR1
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1) ... (24)
US-09-821-616-9
```

Query Match 100.0%; Score 2771; DB 4; Length 534;

Best Local Similarity 100.0%; Pred. No. 1.1e-231; Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 MSFRSLALSGVCTGLANVSKRATLDSMLSEATVARTALINNGADGAWSGADSGI 60
Db 1 MSFRSLALSGVCTGLANVSKRATLDSMLSEATVARTALINNGADGAWSGADSGI 60
Qy 61 VVASPSTNDPDYFTYTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIVOGINSNPSG 120
Db 61 VVASPSTNDPDYFTYTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIVOGINSNPSG 120
Qy 121 DLSGAGLGEPRKFNVDATATGSGWRPQRDPALRATAMIGFGQWLLDNGYSTATDIW 180
Db 121 DLSGAGLGEPRKFNVDATATGSGWRPQRDPALRATAMIGFGQWLLDNGYSTATDIW 180
Qy 121 DLSGAGLGEPRKFNVDATATGSGWRPQRDPALRATAMIGFGQWLLDNGYSTATDIW 180
Db 121 DLSGAGLGEPRKFNVDATATGSGWRPQRDPALRATAMIGFGQWLLDNGYSTATDIW 180
Qy 181 PLVRNDLSVYAOYWNQGYDLMEEVNSSFPTIAVQHRALVEGSARFATAVGSSCWCDSQ 240
Db 181 PLVRNDLSVYAOYWNQGYDLMEEVNSSFPTIAVQHRALVEGSARFATAVGSSCWCDSQ 240
Qy 241 APEILCYLQSFMTGSFTLANFDSRSKGDANTLLGSIHTFDPACDSTFOPCSPRALA 300
Db 241 APEILCYLQSFMTGSFTLANFDSRSKGDANTLLGSIHTFDPACDSTFOPCSPRALA 300
Qy 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPFCTTAAAEQLDALYQMD 360
Db 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPFCTTAAAEQLDALYQMD 360
Qy 361 KQGSLEVTVDVSLDFFKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFPISIVETHAASNG 420
Db 361 KQGSLEVTVDVSLDFFKALYSDAATGTYSSTSSSTYSIVDAVKTFADGFPISIVETHAASNG 420
Qy 421 SMEQYDKSDGEOLSARDLTWSYALLTANRRNSVVPASWGETSASVPGTCAATSAIG 480
Db 421 SMEQYDKSDGEOLSARDLTWSYALLTANRRNSVVPASWGETSASVPGTCAATSAIG 480
Qy 481 TYSSVTVTSPSPSVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTRSGMSL 534
Db 481 TYSSVTVTSPSPSVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTRSGMSL 534
```

## RESULT 7

US-09-351-814-13  
 ; Sequence 13, Application US/09351814  
 ; Patent No. 6352851  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Nielsen, Bjarne Roenfeldt  
 ; APPLICANT: Nielsen, Allan  
 ; APPLICANT: Svendsen, Henrik  
 ; APPLICANT: Pedersen, Henrik  
 ; APPLICANT: Vind, Jesper  
 ; APPLICANT: Hendiksen, Hanne Vang  
 ; APPLICANT: Frandsen, Torben Peter  
 ; TITLE OF INVENTION: Glucoamylase Variance  
 ; FILE REFERENCE: 5636-200-US  
 ; CURRENT APPLICATION NUMBER: US/09351, 814  
 ; CURRENT FILING DATE: 1999-07-12  
 ; EARLIER APPLICATION NUMBER: PA 1998 00937  
 ; EARLIER FILING DATE: 1998-07-15  
 ; EARLIER APPLICATION NUMBER: PA 1998 01667  
 ; EARLIER FILING DATE: 1998-12-17  
 ; EARLIER APPLICATION NUMBER: 60/093, 528  
 ; EARLIER FILING DATE: 1998-07-21  
 ; EARLIER APPLICATION NUMBER: 60/115, 545  
 ; EARLIER FILING DATE: 1999-01-12  
 ; NUMBER OF SEQ ID NOS: 81  
 ; SOFTWARE: FastrSeq for Windows Version 3.0  
 ; SEQ ID NO 13  
 ; LENGTH: 640  
 ; TYPE: PR1  
 ; ORGANISM: ASPERGILLUS NIGER  
 ; US-09-351-814-13

Query Match 99.0%; Score 2742; DB 4; Length 640;  
 Best Local Similarity 99.6%; Pred. No. 4, 9e-229;  
 Matches 528; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSFRLSLALSGVCTGGLANVYSKRATLDSMLNEATVAPRAILNNIGADGAWYSGADSGI 60  
 DB 1 MSFRLSLALSGVCTGGLANVYSKRATLDSMLNEATVAPRAILNNIGADGAWYSGADSGI 60  
 QY 61 VVASPSTNDPDEYFTWTRDSGLVKTLYDLFRNGDTSLSSTIENYISAQAVOGISNPSG 120  
 DB 61 VVASPSTNDPDEYFTWTRDSGLVKTLYDLFRNGDTSLSSTIENYISAQAVOGISNPSG 120  
 QY 121 DISSGAGLGEPRKNNVDETATYTSWGRPQRDGPALRATAMIFGQWLLDNGYSTATDIYW 180  
 DB 121 DISSGAGLGEPRKNNVDETATYTSWGRPQRDGPALRATAMIFGQWLLDNGYSTATDIYW 180  
 QY 181 PLVRNDLSVVAQVWNGQYDLMEVNGSSFFETIAVQHRALVEGSAFATAVSSCSWCDQ 240  
 DB 181 PLVRNDLSVVAQVWNGQYDLMEVNGSSFFETIAVQHRALVEGSAFATAVSSCSWCDQ 240  
 QY 241 APEILCYLQSFMTGSLANFDSRSRSGKDNATLGSITHTDPAAACDSTFQCPSPALA 300  
 DB 241 APEILCYLQSFMTGSLANFDSRSRSGKDNATLGSITHTDPAAACDSTFQCPSPALA 300  
 QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABQLYDALYQMD 360  
 DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABQLYDALYQMD 360  
 QY 361 KQSSLEVTVDLDFPALYSDAATGYSSSSSTYSSTVDAVKTFADGFWISVETHAASNG 420  
 DB 361 KQSSLEVTVDLDFPALYSDAATGYSSSSSTYSSTVDAVKTFADGFWISVETHAASNG 420  
 QY 421 SMSEQYDKSDGDELSARDLTMSYAALLTANRRNSVVPASMGESASVVGTCATSAIG 480  
 DB 421 SMSEQYDKSDGDELSARDLTMSYAALLTANRRNSVVPASMGESASVVGTCATSAIG 480  
 QY 481 TYSSTVTSWPSIVATGGTTTATPTGSGSVTSTSKTATASKTSTTRS 530  
 DB 481 TYSSTVTSWPSIVATGGTTTATPTGSGSVTSTSKTATASKTSTTRS 530

## RESULT 8

US-09-236-063-1  
 ; Sequence 1, Application US/09236063  
 ; Patent No. 6537792  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Allen, Martin  
 ; APPLICANT: Fang, Tsuei-Yun  
 ; APPLICANT: Li, Yuxing  
 ; APPLICANT: Liu, Hsuan-Liang  
 ; APPLICANT: Chen, Hsui-Mei  
 ; APPLICANT: Coutinho, Pedro  
 ; APPLICANT: Hanzacko, Richard  
 ; APPLICANT: Ford, Clark  
 ; TITLE OF INVENTION: PROTEIN ENGINEERING OF GLUCOAMYLASE TO  
 ; TITLE OF INVENTION: INCREASE PH OPTIMUM, SUBSTRATE SPECIFICITY AND  
 ; NUMBER OF SEQUENCES: 12  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Kohn & Associates  
 ; STREET: 30500 No. 653792thwestern Hwy.  
 ; CITY: Farmington Hills  
 ; STATE: Michigan  
 ; COUNTRY: US  
 ; ZIP: 48334  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09236, 063  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Kohn, Kenneth I.  
 ; REGISTRATION NUMBER: 30,955  
 ; REFERENCE/DOCKET NUMBER: 0812.00001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (248) 539-5050  
 ; TELEFAX: (248) 539-5055  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 616 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHEICAL: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Aspergillus  
 ; US-09-236-063-1

Query Match 94.8%; Score 2628; DB 4; Length 616;  
 Best Local Similarity 99.6%; Pred. No. 3, 5e-219;  
 Matches 504; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 25 ATLDLSNEATVAPRAILNNIGADGAWYSGADSGIVVASPSTNDPDEYFTWTRDSGLV 84  
 DB 1 ATLDLSNEATVAPRAILNNIGADGAWYSGADSGIVVASPSTNDPDEYFTWTRDSGLV 84  
 QY 85 KTLVDFRNGDTSLSSTIENYISAQAVOGISNPSGDLSSGAGLGEPRKNNVDETATYTSW 144  
 DB 61 KTLVDFRNGDTSLSSTIENYISAQAVOGISNPSGDLSSGAGLGEPRKNNVDETATYTSW 120  
 QY 145 GRPQDGPALRATAMIFGQWLLDNGYSTATDIWPLVRNDLSVVAQVWNGQYDLME 204  
 DB 121 GRPQDGPALRATAMIFGQWLLDNGYSTATDIWPLVRNDLSVVAQVWNGQYDLME 180  
 QY 205 VNGSSFFETIAVQHRALVEGSAFATAVSSCSWCDQAPETILCYLQSFMTGSLANFDS 264  
 DB 181 VNGSSFFETIAVQHRALVEGSAFATAVSSCSWCDQAPETILCYLQSFMTGSLANFDS 240

Qy	265	RSKCDANTLLGSIHETFDPEAACDDSDTPOCSPRALANKEVVDSPRSIYTLNDGLSDSEA	334
Db	241	RSKCDANTLLGSIHETFDPEAACDDSDTPOCSPRALANKEVVDSPRSIYTLNDGLSDSEA	300
Qy	325	VAVGRIPEDTYYNGNEMFLCTLAALAEOLYDALYOMDKGSLSEVTDVSLDFFKALYSDAAT	384
Db	301	VAVGRIPEDTYYNGNEMFLCTLAALAEOLYDALYOMDKGSLSEVTDVSLDFFKALYSDAAT	360
Qy	385	GTWSSSSSYTSSLYDAVKTFADGFAFSIYETHAASNGSSEQYDSDGSEQLSARDLTMSYA	444
Db	361	GTWSSSSSYTSSLYDAVKTFADGFAFSIYETHAASNGSSEQYDSDGSEQLSARDLTMSYA	420
Qy	445	ALLTANNRRNSVVPASWGETSASVPGTCAATSAIGTSSVTVTWSPIVATGGTITTAAT	504
Db	421	ALLTANNRRNSVVPASWGETSASVPGTCAATSAIGTSSVTVTWSPIVATGGTITTAAT	480
Qy	505	PTGSGSVTSTKTTATASTKSTTTTSS	530
Db	481	PTGSGSVTSTKTTATASTKSTTTTSS	506

```

RESULT 9
US-09-199-290-34
; Sequence 34, Application US/09199290
; Patent No. 6255084
; GENERAL INFORMATION:
; APPLICANT: Nielsen, Bjarne R.
; APPLICANT: Nielsen, Ruby
; APPLICANT: Lehmbeck, Jan
; TITLE OF INVENTION: Thermostable Glucoamylase
; FILE REFERENCE: 5279,200-US
; CURRENT APPLICATION NUMBER: US/09/199,290
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1557/97
; EARLIER FILING DATE: 1997-12-30
; EARLIER APPLICATION NUMBER: 0925/98
; EARLIER FILING DATE: 1998-07-10
; EARLIER APPLICATION NUMBER: 60/070,746
; EARLIER FILING DATE: 1998-01-08
; EARLIER APPLICATION NUMBER: 60/094,344
; EARLIER FILING DATE: 1998-07-28
; EARLIER APPLICATION NUMBER: 08/979,673
; EARLIER FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: 09/107,657
; EARLIER FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 618
; TYPE: PRP
; ORGANISM: Talaromyces emersonii
; US-09-199-290-34

```

	Query Match	Similarity	61.0%;	Score 1689;	DB 3;	Length 618;
	Best Local	Similarity	61.2%;	Pred. No. 9.1e-138;		
	Matches	322;	Conservative	75;	Mismatches	111;
					Indels	18;
					Gaps	5;
QY	7	LALSGLVCTGIANT-ISKRAF-LDSWLSNEATYARIALINNIGADGAVSGADSGIYVA	63			
Db	9	LCIIIGLPAAAPARA.PVAAARATGSLDSFLATETPTIAQVLNNIGNGADVAGAGIYVA	68			
QY	64	SPSTDNPYPFTWTRDGLVTKTYLDLPRNGDTSLSTENYISAOALVQGISNPSGDIS	123			
Db	69	SPSNDPEYFISWTRDPAALTKYLVDAFIAAGKDLKQITLQOYISAOARVQITISNPSGDIS	128			
QY	124	SGAGLGEKENVDEYATYGSWGR.PORDGPALRATIMIGFGOMLDNGYSTATDIWVPLV	183			
Db	129	TG-GLGEKEFVNNEETAFGPMGR.PORDGPALRATIALIYAVYLLIDNGEASTADEIIPYIV	187			
QY	184	RNDLSIYVAQVNNQCGYDLMEEVNCSFFTIVQHQALVEGSAFAFVAGSSCWSDOAPE	243			
Db	188	QNDLSIYITQVNNSSFTFDLMEEVEGSSFTTIVQHQALVEGNAALRLNHTCSNCSQADPO	247			

```

QY      244  IICVLOSFMPTGFEIILANF-DSSRSKQDANLTLGSIHFEDBEAACDDSTFQCSBRLAHN 3020
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      248  VUCFLOSVMYTGSIYVLANFGSGRSKQDVNLSIGSIHFDPDAGGDDSTFQCSARLALAHN 3070

QY      303  KEVWDSFRSIIYLANDGLSDSEAVANGRYPEDDTYNNGNPMFLCTIAAEQYDALYQWCKQ 362
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      308  KVVTSDFRSIIYVANGSIGIAEGSAVAAGRPEDEVYGGNGPMYIATATAAEQYDALYQWKI 367

QY      363  GSLEIVDVSLDFPFKALYSDPATGTSSSSSYTSIYDAVAKTFADGAFSIIYETHAASNGSM 422
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      368  GSIISITDVSLPFQDIIYPSAANGTYNCSGTFPNDIIASVQYIGSGIYSLYENKTFPSDGL 427

QY      423  SEQYDKSDGEQSLARDLTIWSYALITLANNRNSVVPASWGETSASSYPTGCAATSALGTY 482
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      428  TEOFSTRDTPISASALTIWSYASLITTAARQSVVPASWGSSASVPAVCASATSATGPY 487

QY      483  SSVTVYMSPIVATGGTITTAAPTQSGVSTSKTATASKTSTTT 528
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      488  STAKINTWPS-----SSGSGSITTSAPCTTPPSAVT 520

```

```

RESULT 10
US-09-821-616-34
Sequence 34. Application US/09821616
Patent No. 6620924
GENERAL INFORMATION:
APPLICANT: Nielsen, Bjarne R.
APPLICANT: Nielsen, Ruby
APPLICANT: Lehmebeck, Jan
TITLE OF INVENTION: Thermostable Glucoamylase
FILE REFERENCE: 5279,200-US
CURRENT APPLICATION NUMBER: US/09/821,616
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/199,290
PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-24
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 1557/97
PRIOR FILING DATE: EARLIER FILING DATE: 1997-12-30
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 0925/98
PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-10
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/070,746
PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-08
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/094,344
PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-28
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/979,673
PRIOR FILING DATE: EARLIER FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/107,657
PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-30
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 34
LENGTH: 618
TYPE: BPT

```

[illegible]

```

Db 188 QNDLSITTOYWNSTEDLMEVEGSSFFTTAVOHRALVBSGNALATRLNHTCSNCVSOAPQ 247
Qy 244 ILCYLOSFWTGSFILANF-DSSRSGKDNANTLGSIHTEPEAACDDSTFQPCSPRALANH 302
Db 248 VLFCLDSYMTGSLVLANFGSGSGKDVNSILGSIHTFDPAGCCDDSTFQPCSPRALANH 307
Qy 303 KEVVDSPRSITLINDGLSDSEAVAVGRYPEDITYNGNPMFLCTLAABEOLYDALYOMDKO 362
Db 308 KVTJDSFRSITVAINSGIAEGSAVAVGRYPEDIVYQGGNPMFLATYAAAEQLYDAIYOMKXI 367
Qy 363 GSELEVDVSLDEPKALYSOAATCTSSSSSTSYIDAVKTFADGFSVIVETHAANGSM 422
Db 368 GSISITVDLPPFQODIYPSAAVGTYNSTGTFNDIISAVQTYDGYLSIVKXTYPSDGL 427
Qy 423 SEQYDSDSEQLSARPLTWSYALLTANNRRNSVVPASMGTSASVPGTCAATSAIGTY 482
Db 428 TEPFSITDGPPLSASALTSYASLITRSAROSVVPASMGESSASVPAVCATSATGPY 487
Qy 483 SSVTVTSWPSIVATGTTTATPTGSGSVTSKTTATASKTSTTT 528
Db 488 STATWTWPS-----SGSGSSTTSSAPCTTPTSVAVT 520

```

```

RESULT 11
US-09-199-290-7
/ Sequence 7, Application US/09199290
/ Patent No. 6255084
/ GENERAL INFORMATION:
/ APPLICANT: Nielsen, Bjarne R.
/ APPLICANT: Nielsen, Rudy
/ APPLICANT: Lehmebeck, Jan
/ TITLE OF INVENTION: Thermostable Glucoamylase
/ FILE REFERENCE: 5279.200-US
/ CURRENT APPLICATION NUMBER: US/09/199,290
/ CURRENT FILING DATE: 1998-11-24
/ EARLIER APPLICATION NUMBER: 1557/97
/ EARLIER FILING DATE: 1997-12-30
/ EARLIER APPLICATION NUMBER: 0925/98
/ EARLIER FILING DATE: 1998-07-10
/ EARLIER APPLICATION NUMBER: 60/070,746
/ EARLIER FILING DATE: 1998-01-08
/ EARLIER APPLICATION NUMBER: 60/094,344
/ EARLIER FILING DATE: 1998-07-28
/ EARLIER APPLICATION NUMBER: 08/979,673
/ EARLIER FILING DATE: 1997-11-26
/ EARLIER APPLICATION NUMBER: 09/107,657
/ EARLIER FILING DATE: 1998-06-30
/ NUMBER OF SEQ ID NOS: 34
/ SOFTWARE: PaetsEQ for Windows Version 3.0
/ SEQ ID NO 7
/ LENGTH: 591
/ TYPE: PRT
/ ORGANISM: Talaromyces emersonii
US-09-199-290-7

```

```

Query Match 60.4%; Score 1674.5; DB 3; Length 591;
Best Local Similarity 62.3%; Pred. No. 1.5e-136;
Matches 314; Conservative 73; Mismatches 102; Indels 15; Gaps 3;

```

```

Qy 26 TLDWSLNEATVARTAILNNGIGADGAMVSGADSGIVASPSNDPPIFYTWTRDGLVTK 85
Db 4 SLDSPFLATEPPLAQVLNNGIGADGAMVSGADSGIVASPSRSDPPIFYTWTRDGLVTK 63
Qy 86 TLVDLFRNDTSLSTIENTENISAOAVOGISNPSGDISGAGIGEPKFNVDETAYTGSWG 145
Db 64 YLVDAFNRGKMLEQITIQYISAQAKVQITSNPSGDISG-GLGEPKFNVDETAYTGSWG 122
Qy 146 RPDGDPALRATAMIGFGWLNDNGYTSTATDIWPLVENDLSYVAQYNNQGTGYDLMEVY 205
Db 123 RPDGDPALRATAMIGFGWLNDNGYTSTATDIWPLVENDLSYVAQYNNQGTGYDLMEVY 182
Qy 206 NGSSFTTIAVQHRALVEGSAFATAVSGSCGCDGSAPEILCYLOSFWTGSFILANF-DSS 264

```

```

Db 183 EGSSFTTIAVQHRALVEGSAFATAVSGSCGCDGSAPEILCYLOSFWTGSFILANFGSG 242
Qy 245 RSGKDNANTLGSIHTEPEAACDDSTFQPCSPRALANHKEVVDSPRSITLINDGLSDSEA 324
Db 243 RSGKDNANTLGSIHTEPEAACDDSTFQPCSPRALANHKEVVDSPRSITLINDGLSDSEA 302
Qy 325 VAVGRYPEDITYNGNPMFLCTLAABEOLYDALYOMDKOSLEVTDVSLDFKALYSDAAT 364
Db 303 VAVGRYPEDIVYQGGNPMFLATYAAAEQLYDAIYOMKXISITDVLPPFQODIYPSAAV 362
Qy 365 GTSYSSSTSYSSIVDAVKTFADGFSVIVETHAANGSMSEBQYKXSGEQLSARPLTWSYA 444
Db 363 GTSYSSSTTFNDIISAVQTYDGYLSIVKXTYPSDGLTEQFSRTGTPLSASALTWSSYA 422
Qy 445 ALLTANNRRNSVVPASMGTSASVPGTCAATSAIGTSVTVTSWPSIVATGTTTAT 504
Db 423 SLTTAARQOSVVPASMGESSASVLAVCATSATGPYSTATTWTWPS----- 470
Qy 505 PTGSGSVTSKTTATASKTSTTT 528
Db 471 -SGSGSSTTSSAPCTTPTSVAVT 493

```

```

RESULT 12
US-09-821-616-7
/ Sequence 7, Application US/09821616
/ Patent No. 6620924
/ GENERAL INFORMATION:
/ APPLICANT: Nielsen, Bjarne R.
/ APPLICANT: Nielsen, Rudy
/ APPLICANT: Lehmebeck, Jan
/ TITLE OF INVENTION: Thermostable Glucoamylase
/ FILE REFERENCE: 5279.200-US
/ CURRENT APPLICATION NUMBER: US/09/821,616
/ CURRENT FILING DATE: 2001-03-29
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/199,290
/ PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-24
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 1557/97
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-12-30
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 0925/98
/ PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-10
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/070,746
/ PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-08
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/094,344
/ PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-28
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/979,673
/ PRIOR FILING DATE: EARLIER FILING DATE: 1997-11-26
/ PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/107,657
/ PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-30
/ NUMBER OF SEQ ID NOS: 34
/ SOFTWARE: PaetsEQ for Windows Version 3.0
/ SEQ ID NO 7
/ LENGTH: 591
/ TYPE: PRT
/ ORGANISM: Talaromyces emersonii
US-09-821-616-7

```

```

Query Match 60.4%; Score 1674.5; DB 4; Length 591;
Best Local Similarity 62.3%; Pred. No. 1.5e-136;
Matches 314; Conservative 73; Mismatches 102; Indels 15; Gaps 3;

```

```

Qy 26 TLDWSLNEATVARTAILNNGIGADGAMVSGADSGIVASPSNDPPIFYTWTRDGLVTK 85
Db 4 SLDSPFLATEPPLAQVLNNGIGADGAMVSGADSGIVASPSRSDPPIFYTWTRDGLVTK 63
Qy 86 TLVDLFRNDTSLSTIENTENISAOAVOGISNPSGDISGAGIGEPKFNVDETAYTGSWG 145
Db 64 YLVDAFNRGKMLEQITIQYISAQAKVQITSNPSGDISG-GLGEPKFNVDETAYTGSWG 122
Qy 146 RPDGDPALRATAMIGFGWLNDNGYTSTATDIWPLVENDLSYVAQYNNQGTGYDLMEVY 205
Db 123 RPDGDPALRATAMIGFGWLNDNGYTSTATDIWPLVENDLSYVAQYNNQGTGYDLMEVY 182

```

```

QY      200  NSSPFTLIVORALVEGSAFTATVAGSSCMSCDSDAPETLICYLOSFWTGSIIANF--DSS 264
      |||||
Db      183  EGSSPFTTIVQORALVEGNALATRLNHTCSNCVSQAPVULFLOSITGSGVILANFGSG 242

QY      265  RSGKDANTLLGSIHTFDEPAACDSDTFOPCSPRALANHEKVEVDSFRSIYTLINDGLSDSEA 324
      |||||
Db      243  RSGKDVNSTLGSIHFTDPAGCGDSDTFQPCSRBALANHKVYTDSEFRSIYALNNGSLAGSA 302

QY      325  VAVGRYPEDTYNGNPMFLCTIAAAEQYDALYOMKOGSLEVTDSLDEPFKALYSDAAT 384
      |||||
Db      303  VAVGRYPEDVYQGNPNMYLATIAAAEQYDALYOMKKGISITDVSLEPFQDIYPSAAV 362

QY      385  GYSSSSSTSYSSIVDAVKTFADGFPYSIYETHAASNGSSEQYDSDDEQSLARPLTYSYA 444
      |||||
Db      363  GYNSGSTTFENDIISAVQTYGQGVYSIYEKTPSPSGSLTEQFSRSTDDGPLLSASALTWSYA 422

QY      445  ALLTANNRNSVPAASWGEITSASSIPGCAATSAIGYSSSYTVVSWPDSIYVATGTTTTAT 504
      |||||
Db      423  SLLTAAARQSVVPAASWGEISSASSVLAVCASATSAIGPSTRTNTRWPS----- 470

QY      505  PTGSGVTSNKTATATASKTSTTT 528
      |||||
Db      471  -SGSGSTTSSACPCTPTPSVAVT 493

```

```

RESULT 13
US-09-704-449-2
; Sequence 2, Application US/0970444-9
; Patent No. 6309872
; GENERAL INFORMATION:
; APPLICANT: Key, Michael W.
; APPLICANT: Goldightly, Elizabeth J.
; TITLE OF INVENTION: Polypeptides Having Glucosylase
; TITLE OF INVENTION: Activity And Nucleic Acids Encoding Same
; FILE REFERENCE: 10101, 000-US
; CURRENT APPLICATION NUMBER: US/09/704, 449
; CURRENT FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 630
; TYPE: .PRT
; ORGANISM: Thielavia terrestris
; US-09-704-449-2

```

Query Match	58.3%;	Score 1615.5;	DB 4;	Length 630;
Best Local Similarity	58.9%;	Pred. No. 2.2e-131;		
Matches 315;	Conservative 83;	Mismatches 112;	Indels 25;	Gaps 7;

QY	6	ILALSGVCTGLANNVTSKKA-----TLDSTWLSNATVARTILINNIIGADAWGSA	56
Db	6	ILGLALLPALGHPASRVRREGEVAKRSVDSFIATESPIALSNULCNIGSTCHASGV	65
QY	57	DGSIIVASPBSTNDPVEFYTWTRDSGLVYKTYLVDLFRNG-DPLUSTJENYSQAIVOGI	115
Db	66	ASGIIVASPEKTNPDWYTWYTRDSALTFECVVDITNSYDLSLQEIONYIWAAPHIQGV	123
QY	116	SNBSGDLSSGAGLGEPKFNVDETAAYTGSNGRPQDPGRALRTAMIGFCQMLLDNGYSTA	175
Db	126	SNPSGSLSDSGEGEPKFNVDMSGFGAAGRPOQRGPPALRALIAYSKWILSNGYSTA	185
QY	176	TDIYVWLVANNDLSYVAVQYNNQTCGYLMEWVNSSPFTLANQHRALVESSAFATVSSCS	235
Db	186	SSLIYVWLVATKIDLAVYVQ--NNTGFPDLMEEVSSSSFTYVANOHRALVESSALATSLGTSCS	243
QY	236	WCDSQAPELICYLQGSFWTGS--FIILANFDSRSRGKDANTLLGSIHTPDEAACDSTFOP	293
Db	244	ACBAVAPOLICFLQGSFWSPPSSGYLIAN-----STAKDANTLLGSIHTPDEPAAGCDATFOP	299
QY	294	CSGRALANKHEVVDPSRSTIYTLNDGLDSEAAVAVGRYEDPQYNNQNPFIQLTAAAOYL	351
Db	300	CSGRALANKHVYVDARSTIYTSINGSLGSSAAVAVGRYEDPSIFGQNPFIYLTLLAAAOYL	355

```

QY      35 DALYQWDDKQGSLETVDLVLDLFFKALYSPAAGTSSSSSTSSVYDVAKTFPAAGFMSIYE 413
Db      360 DALYWKQKQGSITVTSTSLAFKQDSSSIITGTGYSSTSTYTLTYNAISAYDQGMNIVA 419

QY      414 THAASNGSMSEYDKSDGEOLSAIDLITMSYALITANNRRNSVVPASGELTSASVPGTC 473
Db      420 QYAQNTGSLSEPFKSTNEPLSAIDLITMSYALFLTAARARAGVPPMGGAASANSVPAQC 479

QY      474 AATSAGTSSYTVTMSPSIATAGGTTTATPTGSGVTSKTTATASKTISTTT 528
Db      480 SATSVGGSITSAITSFPP-----SQTPASTSAGS-SPASSTTATATACSTPT 527

```

RESULT 14  
 US-08-596-300A-7  
 Sequence 7, Application US/08596300A  
 Patent No. 5834191  
 GENERAL INFORMATION:  
 APPLICANT:  
 TITLE OF INVENTION: Production of Heterologous Peptides  
 NUMBER OF SEQUENCES: 14  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Klauber & Jackson  
 STREET: 411 Hackensack Avenue  
 CITY: Hackensack  
 STATE: New Jersey  
 COUNTRY: USA  
 ZIP: 07601  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)  
 CURRENT APPLICATION DATA: /08/050 2003

1 FILING DATE: 13-FEB-1996  
2 CLASSIFICATION: 435  
3 ATTORNEY/AGENT INFORMATION:  
4 NAME: Jackson Esq., David A.  
5 REGISTRATION NUMBER: 26,742  
6 REFERENCE/DOCKET NUMBER: 1331-1-0001  
7 TELECOMMUNICATION INFORMATION:  
8 TELEPHONE: 201 487-5800  
9 TELEFAX: 201 343-1684  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
82

Query Match	56.9%;	Score 1577;	DB 2;	Length 626;
Best Local Similarity	57.6%;	Pred. No. 4.8e-128;		
Matches 306; Conservative	85;	Mismatches 130;	Indels 10;	Gaps 6;

[illegible]

Db 192 PIYKNDLATTAYQYNNNTGFDLMEVNNSSFFILASHALVEGSAFKAIVSSGSSCADAI 251

QY 241 APEILCYLOSFWTGS-FILANFDSSRSGKADNTLLGSIHTPEPAACDSTPOECSPRAU 299

Db 252 APQILCFQOOSFWMSNGYIISFNVRNSGKDINSVLTSIHNFDPAAGCQVNTFQCSDRAL 311

QY 300 ANHKEVVDSPRSLIYTLNDLSDSEAVNAGRYPEDTYNGNFWELCTLAARQLYDALYOM 359

Db 312 ANHKVAVDSDNR-FWGNVSGRTAGKAAAGRAVEDYYNGNFWYLTATLAAAOQLDAYVW 370

QY 360 DKQGSLEVTVDLSHDFPKALYSDAATGYSSSSSTYSIVDAKVFADQFVSIIVETHAASN 419

Db 371 KKQGSITVITSTLAFKDLVPSVSGITGYSSSSSTYTAIINAVTYADQFVDIYAQYTPSD 430

QY 420 GSNSEYDYSDEQULSARDLWTSYAAALLTANNRNSVVPASWGETSASVPGTCAATSAI 479

Db 431 GSIABEOPFDKDXGAPLSTATHLWTSYASFLSPAAARRAGIIVPSPWGAASANSILGSCSASTVA 490

QY 480 GTYSALTVSWPBIVAVTGGTTTT-ATPPGSGS-----TYSKTTAATAASKI 524

Db 491 GSATATATATSPANLTPASTVITPTPTQIGCAADHEVLVTEKEKTYTSGQT 541

```

RESULT 15
US-08-596-300A-14
; Sequence 14, Application US/08596300A
; Patent No. 5834191
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Production of Heterologous Peptides
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauder & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/596.300A
FILING DATE: 13-FEB-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1321-I-001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 626 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEtical: NO
FRAGMENT TYPE:
;
US-08-596-300A-14
Query Match 56.9%; Score 1577; DB 2; Length 626;
Best Local Similarity 57.6%; Pred. No. 4,8e-128;
Matches 306; Conservative 85; Mismatches 130; Indels 10; Gaps 6

```

Db	13	AFQAVLGLPDLPEHKRHSIIKR-SVDSYIQTEPRIAQKULCNIGASGRASGAA	71
Qy	62	VASPESTDNPDYIYTYWTRDSGLVATKLVDFPNNG-DTSLSTIENYISAOAI	120
Db	72	VASPSKSSPDYIYTYWTRDAALVTKLIYDFENNDYNTLQNTIQAYAAAKLQGV	131
Qy	121	DISSGAGLPEPKFENVDETAYTGSWCRPORDPALRATMIIGQWMLDNGY	180
Db	132	SLSNAGGLGEPEKFMWDLQOFTGAMKRPORDEPLRALALICYGKMLVNSG	193
Qy	181	PLVRNDLSVAAQYNNQOTGYDLMEEVNGSSFITIAVQHRALVEGSAFAT	240
Db	192	PIYKNDLATAAYANNNTGFDLMEEVNNGSSFITIAASHRALVEGSAFAK	251
Qy	241	APRILCYLGSFPMYGS-FILANPDSRSKQDANTLLGSIHTFDPACDDST	299
Db	252	APRILCYLGSFPMYGS-FILANPDSRSKQDANTLLGSIHTFDPACDDST	311
Qy	300	ANHKVEVDSFRSIYTLNBDGLSDSEAVAVGRYPEDTYNGNFWELCTLA	359
Db	312	ANHKVAVVDSMR-FWGVNSGRTGKAAAGRAEVDYVYNGNFWELCTLA	370
Qy	360	DKQGSLEVTVDLDFPFKALYSDAAGTSSSSSTYSIVAVKTFEADGAF	419
Db	371	KKQGSITVTSTLAFPKLVPSVIGTIGSSSSSTYTALINAVTTIADG	430
Qy	420	GSMSEQYKSDSEOLSARDLTWSYAAALLTANNRRNSVVPASWGETSAS	479
Db	431	GSLAEQFDKDSGAPISATHLITWSYASFLSAARAGIPEPMGAASANS	490
Qy	480	GTYSSVTYMSPSIYATGGTTTT-ATPFGSSS-----VTISKTKTAN	524
Db	491	GSATATATATSPANLTPASITVTPPTQGCADAEHLVLTENEKTYTSG	541

Qy	Db	Query Match	Best local Similarity	Matches	Conservative	Indels	Gaps
Qy	Db	9 LSGLVGTGL--ANVT---SKRATLDWSLSENAETAVARATLAINNTGADGAWYSGADSGIVVA 63	51.4%; Score 1425.5; DB 4; Length 581;	273;	81;	146;	23;
Qy	Db	6 LYLGLVASALMGQGVVAPSPKDNLSERFTDKQADISIKGLVANIGADGRKQAGAPGAIVA 65	52.2%; Pred. No. 5.8e-115;	273;	81;	146;	23;
Qy	Db	64 SPSTNDPDIYFTWTRDSGLVLTIVDLFRNGDTSLSLTNIENYISAQAIYVGISNPSGDS 123		273;	81;	146;	23;
Qy	Db	66 SPEKEDPDYMTWTWTRDSALTYKVLVERFTIHGDKSLQKRIDEYVSAQAKLQGTNNPSGSE 125		273;	81;	146;	23;
Qy	Db	124 SGAAGLGEPEFENVDENATATATGSMGRPPORDPALRATMTMGFGQMLDNGYTSIANDIYWPV 183		273;	81;	146;	23;
Qy	Db	126 SG-GAGEPEFHNULTAFGWSWGRPPORDPPLRATPALTLTYEMWLI SHERSKALNTKWPEVI 184		273;	81;	146;	23;



[illegible]

```

RESULT 17
US-09-999-201B-4
: Sequence 4, Application US/09999201B
: Patent No. 651804
: GENERAL INFORMATION:
: APPLICANT: Berka, Randy
: APPLICANT: Rey, Michael
: APPLICANT: Brown, Kimberly
: TITLE OF INVENTION: Promoters For Expressing Genes In A
: TITLE OF INVENTION: Fungal Cell
: FILE REFERENCE: 5611.210-US
: CURRENT APPLICATION NUMBER: US/09/999,201B
: CURRENT FILING DATE: 2001-10-30
: PRIOR APPLICATION NUMBER: 09/534,407
: PRIOR FILING DATE: 2000-03-22
: PRIOR APPLICATION NUMBER: 09/274,449
: PRIOR FILING DATE: 1999-03-22
: NUMBER OF SEQ ID NOS: 40
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4
: LENGTH: 581
:
: TYPE: PRF
: ORGANISM: Fusarium
: US-09-999-201B-4

```

	Query Match	51.4%	Score 1425.5;	DB 4;	Length 581;
	Best Local Similarity	52.2%	Pred. No. 5.8e-115;		
	Matches	273;	Conservative	81;	Mismatches 146; Indels 23; Gaps 6
QY	9	LSGLVCTGL--ANVI---SKRATLDSMLNEATVARTALINNIGADGAWSGADGIIVA	63		
Db	6	LYGLVASALMGQQVAVASPSKNSLERFIIDKQKDISIKGYLANIGADGRAGQAARGAVA	65		
QY	64	SPTSDNPPIFYTTWTPDSSGLVTKLYLDLRNGDTSLSTIENTYSQAOLVGISNPGCLS	123		
Db	66	SPSKEDPPYTWWTDSALTYKVLERFIHDGKSJQRKIDEVVSQAOKTGTPNSGSPE	125		
QY	124	SGAGLGEEKEFNVDRAAYTGWMCRRPDRDGALRATAMIGCWLITGVTSRTATDIWVPLV	183		
Db	126	SG-GLGEERKEFVNLTAFITGSMCRPRDDGPRLRATALTIAEVLISHGERSKALKINRWVPVT	184		
QY	184	RNDLSYVAQYNQNTGYDLMEEVNGSSFFPIAIOHVALVEGSAFAFAVAGSSCSGCISOAPE	243		
Db	185	EKDLATYTRKFNMRITGYDLMEEVNGSSFLLTSASHRALVEGALLAKKLIGSCPDCTYNNPR	244		
QY	244	ILCYIQSWTGSFILIANFD--SSRGCKDANTLGSIHFPDEPAACDSTFPQCSPALAN	301		

```

Db      245  VLCLFLQFWTGGVYDSDNINVDGRKGLDVNSILSIHFDFDPSKCTDSTFCQCSPLALN   304
Oy      302  HKEYVDSFRSLTYTLNDJLGSDEBAVAVERYPBETTYNGNPWFCTLLAAEQLYDALYQMXK   361
Db      305  HKAVVDSFRSLTYGVKNKGOGKLAAYGYSBEVVYDGNPMWYLATLAAEQLYAAYVQMWK    364
Oy      362  QGSELVTVDSLDFFKALYSDAATGYSSSSSPTYSISIVAVAKTFADGFVSIVETHAASNGS   421
Db      365  LGAIVYDDVSLSFEDDIPIKYXSKGYIAKKTKIKRIIAAKTYADGFAVAVQTTPFKDSS     424
Oy      422  MSEQYDKSDGEQLSARDLTWSYIALLTNNRRNSVPVASWGETSASSVPGTCALTSIAGT     481
Db      425  LAEGFDKSTGAPKSAVHLTWISAALVAATTERDRGIIISWGESSANKPYAVCOAAPACDT     484
Oy      482  -----YSSVTWSMPSIATGATTTA-----PPTGSG 509
Db      485  TTITSVKANVOYSSDOKVYVVGSVTELSMWSPPDGALTLPSSSG 527
```

```

18 RESULT 18
US-08-385-370-2
; Sequence 2, Application US/08385370
; Patent No. 5665585
; GENERAL INFORMATION:
; APPLICANT: Torkkeli, Tuula
; APPLICANT: Joutsjoki, Vesa
; APPLICANT: Torkkeli, Helena
; APPLICANT: Vainio, Arja
; APPLICANT: Pajestrom, Richard
; APPLICANT: Aho, Sirpa
; APPLICANT: Korhola, Matti
; APPLICANT: Nevalainen, Helena
; TITLE OF INVENTION: Production of Glucoamylase P with High
; TITLE OF INVENTION: Limit Dextrinase Activity in Trichoderma
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/385,370
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/104,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cimdala, Michele A.
; REGISTRATION NUMBER: 33,851
; REFERENCE/DOCKET NUMBER: 1050.0270004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; TELEX: 248636 SSK
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 616 amino acids
; TYPE: amino acid
; TOPOLOGY: both
US-08-385-370-2

```

Query Match	47.7%;	Score 1321;	DB 1;	Length 616;
Best Local Similarity	48.6%;	Pred. NO. 7.3e-106;		
Matches 257;	Conservative 88;	Mismatches 150;	Indels 34;	Gaps 6;



QY 14 CTGLANVIS-----KRATLDLSLNEATVARTAILNINIGADGAWVSGADSGIYVA 63  
 DB 8 CAGALSILCSLAIAAPTELKARDLSFISERAIALOGALNNIGPDGSAVPGAGAFVYA 67  
 QY 64 SPSTNDPDYFYTWTRDSGLVLTVDLFRNGDTLSLSTIENTYISAGAIYOGISNPGCD-L 122  
 DB 68 SPKANPDYFYTWTRDSGLVLTVDLFRNGDTLSLSTIENTYISAGAIYOGISNPGCD-L 127  
 QY 123 SSGAGIGEPKFNVDERTAYTGSWGRPORDPALRATMIGFGWLDNGYTSRTATDVIWPL 182  
 DB 128 PDGVGIGEPKFNVDERTAYTGSWGRPORDPALRATMIGFGWLDNGYTSRTATDVIWPL 187  
 QY 183 VRNDLSYVAQYNNQTYDYLMEEVNGSSFTTIAVOHRALVEGSAFATAVSSCSWCDQAP 242  
 DB 188 IANDLSYVQYNNQTYDYLMEEVNGSSFTTIAVOHRALVEGSAFATAVSSCSWCDQAP 246  
 QY 243 EILCYLOSFWTGSFILANF--DSSRSGKADNTLGSITHTDPAACDDSTFQPCSPRALA 300  
 DB 247 EYLCLFQSFNGKVIYISNINNVNNGRTGLDGNSTLGAISTFDIDAYCDSPFLQCHQSOLA 306  
 QY 301 NHKEVVDSPRSITYTLNDGLSDSEAVAVGRYPEDTYNGNWPFLCTLAABOLYDALYOND 360  
 DB 307 NFKVLTDFTRNLTYTINAGIPEGGVAVGRYAEVDYNGNWPFLCTLAABOLYDALYOND 366  
 QY 361 KOGSLEVDVSLDFPALYSDAATGYSS--SSSTYSIVDAVKTFADEFSIVETHAAS 418  
 DB 367 ARHVLTVDETSIAFFEDIYPEVTAREYKSGNANSPQAQIMDAVTAADSVVAAEKTIIPS 426  
 QY 419 NSGMEQYKSDGEQLSARDLTWSYAAALTANNRRNSVVPASGGETSASSVPGTCAATSA 478  
 DB 427 NSLSQFNRDGTPLSAIDLWTMSYAAFTMGQRAGQYPSWSGNSNALPPTTCSASST 486  
 QY 479 IGTYSVTVTSMPSIVATGTTTATPTGSGSVTSTKTATASKTST 527  
 DB 487 PGTY-----TPATAAGAPNVTSCQVSTFINNATT 517

# RESULT 19 US-08-385-370-4

; Sequence 4, Application US/08385370  
 ; Patent No. 5665585  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Torkkeli, Tuula  
 ; APPLICANT: Joutsen, Vesa  
 ; APPLICANT: Torkkeli, Helena  
 ; APPLICANT: Vainio, Arja  
 ; APPLICANT: Fagerstrom, Richard  
 ; APPLICANT: Aho, Sirpa  
 ; APPLICANT: Korhola, Matti  
 ; APPLICANT: Nevalainen, Helena  
 ; TITLE OF INVENTION: Production of Glucoamylase P with High  
 ; TITLE OF INVENTION: Limit Dextrinase Activity in Trichoderma  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
 ; STREET: 1100 New York Ave., Suite 600  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20005  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/385,370  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/104,853  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:

; NAME: Cimbal, Michele A.  
 ; REGISTRATION NUMBER: 33,851  
 ; REFERENCE/DOCKET NUMBER: 1050.0270004  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 371-2600  
 ; TELEFAX: (202) 371-2540  
 ; TELEX: 248636 SSK  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 616 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-385-370-4

Query Match 47.7%; Score 1321; DB 1; Length 616;  
 Best Local Similarity 48.6%; Pred. No. 7,3e-106;  
 Matches 257; Conservative 88; Mismatches 150; Indels 34; Gaps 6;

QY 14 CTGLANVIS-----KRATLDLSLNEATVARTAILNINIGADGAWVSGADSGIYVA 63  
 DB 8 CAGALSILCSLAIAAPTELKARDLSFISERAIALOGALNNIGPDGSAVPGAGAFVYA 67  
 QY 64 SPSTNDPDYFYTWTRDSGLVLTVDLFRNGDTLSLSTIENTYISAGAIYOGISNPGCD-L 122  
 DB 68 SPKANPDYFYTWTRDSGLVLTVDLFRNGDTLSLSTIENTYISAGAIYOGISNPGCD-L 127  
 QY 123 SSGAGIGEPKFNVDERTAYTGSWGRPORDPALRATMIGFGWLDNGYTSRTATDVIWPL 182  
 DB 128 PDGVGIGEPKFNVDERTAYTGSWGRPORDPALRATMIGFGWLDNGYTSRTATDVIWPL 187  
 QY 183 VRNDLSYVAQYNNQTYDYLMEEVNGSSFTTIAVOHRALVEGSAFATAVSSCSWCDQAP 242  
 DB 188 IANDLSYVQYNNQTYDYLMEEVNGSSFTTIAVOHRALVEGSAFATAVSSCSWCDQAP 246  
 QY 243 EILCYLOSFWTGSFILANF--DSSRSGKADNTLGSITHTDPAACDDSTFQPCSPRALA 300  
 DB 247 EYLCLFQSFNGKVIYISNINNVNNGRTGLDGNSTLGAISTFDIDAYCDSPFLQCHQSOLA 306  
 QY 301 NHKEVVDSPRSITYTLNDGLSDSEAVAVGRYPEDTYNGNWPFLCTLAABOLYDALYOND 360  
 DB 307 NFKVLTDFTRNLTYTINAGIPEGGVAVGRYAEVDYNGNWPFLCTLAABOLYDALYOND 366  
 QY 361 KOGSLEVDVSLDFPALYSDAATGYSS--SSSTYSIVDAVKTFADEFSIVETHAAS 418  
 DB 367 ARHVLTVDETSIAFFEDIYPEVTAREYKSGNANSPQAQIMDAVTAADSVVAAEKTIIPS 426  
 QY 419 NSGMEQYKSDGEQLSARDLTWSYAAALTANNRRNSVVPASGGETSASSVPGTCAATSA 478  
 DB 427 NSLSQFNRDGTPLSAIDLWTMSYAAFTMGQRAGQYPSWSGNSNALPPTTCSASST 486  
 QY 479 IGTYSVTVTSMPSIVATGTTTATPTGSGSVTSTKTATASKTST 527  
 DB 487 PGTY-----TPATAAGAPNVTSCQVSTFINNATT 517

## RESULT 20

; Sequence 11, Application US/08270076A  
 ; Patent No. 5667966  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sleep, Darrell  
 ; APPLICANT: Goodey, Andrew R  
 ; APPLICANT: Vakeria, Diana  
 ; TITLE OF INVENTION: Yeast Promoter  
 ; NUMBER OF SEQUENCES: 17  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: The BOC Group, Inc.  
 ; STREET: 100 Mountain Avenue, Murray Hill  
 ; CITY: New Providence  
 ; STATE: New Jersey  
 ; COUNTRY: USA  
 ; ZIP: 07974

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/270,076A
FILING DATE: 01-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 8923521.2
FILING DATE: 18-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/597,687
FILING DATE: 16-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/925,286
FILING DATE: 04-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Swope, R. Hain
REGISTRATION NUMBER: 24864
REFERENCE/DOCKET NUMBER: 92H834-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 908/771-6292
TELEFAX: 908/771-6159
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 806 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-270-076A-11
```

```
Query Match 21.3%; Score 589; DB 1; Length 806;
Best Local Similarity 30.6%; Pred. No. 3,36-42;
Matches 15; Conservative 90; Mismatches 200; Indels 60; Gaps 15;
```

```
QY 1 MSFRLSLALSGLVCTGLANVSKRATLDSWLSNEATVAFATLAINIGAGAWVSGADSGI 60
D 13 VQLEADVIANMGTVYVD-SNGAMDSALBEMLQROKAVSIERIFENIGSAVYPS-ILPGV 370
QY 61 VVASPSTNDNPDYFTYTRDLSGLVKTIVDLFRNDTSLSTIETIYIAQALVQGISNPSG 120
D 371 VIAPSPQTHDPYFQWIRDSALRTINSIVS--HSADPA-IETLQYLVNVSFHLQRTNN--- 424
QY 121 DLSSGAG-----LGEPEVNDETAYTGSWGRFQPDGPALRATAMIGFGWILDNG--- 170
D 425 TLGAGIGTYNDYVALGDPKMNVDATPTPEWGRFQNDGPALRSIAILIKIIDYIKQSGTDL 484
QY 171 -----YTSATDIWPLVRNDLSVAAQYMNQGYDLMEVYVNGSSFFTIAYQHRALVGS 225
D 485 GAKYFPQSTA-DIPDDIVRWYLRFIIDHWNSSGFDLMEVGMHFFTLVQLSAVDRTLS 543
QY 226 FATVAGSSCSCMD---SQAPILCYL-----QSFMTGSFILANF-----DSRSQKDA 270
D 544 YFNASERSSPFVELEQRTDRDISKFLVDPAANGFINGKY--NYIVETPMIADTLRSGLDI 600
QY 271 NTLLGSIHTPPEAACDSTFQPCSPRALANHKEVDSFNSIYTLNDGLSDSEAVANGRY 330
D 601 STLIANTVHDAPSA-SHLPFDINDPAVLNTLHLMHMSISYIYNSSKNATGIALGRY 659
QY 331 PEDTY-----YNGNPWFLCTIAAEOLYDALYQMDKQSLSEVTDSIDFFKALYSDATG 385
D 660 PEDYDYGVGVEGHPWLATCASTLYQLIYRISEGHDLVVMNNDCSNAFWSELVFS 719
QY 386 TYSS-----SSSTYSSIVDAVKTPADGFSVIVETHAASNGSMGQYKXSGEQ 433
D 720 NLITLGNDEGYLLEFPTAFNQIYKIFOLADSLFLVKAHVGTGDELSEQFNKYTFM 779
QY 434 LSARDLTMSYAALLTANRRNSV 457
D 780 QGAGHLLTWSTYSFMDAYQIROEVL 803
```

```
RESULT 21
US-09-199-290-4
Sequence 4, Application US/09199290
Patent No. 6255084
GENERAL INFORMATION:
APPLICANT: Nielsen, Bjarne R.
APPLICANT: Nielsen, Rudy
TITLE OF INVENTION: Thermostable Glucoamylase
FILE REFERENCE: 5279,200-US
CURRENT APPLICATION NUMBER: US/09/199,290
CURRENT FILING DATE: 1998-11-24
EARLIER APPLICATION NUMBER: 1557/97
EARLIER FILING DATE: 1997-12-30
EARLIER APPLICATION NUMBER: 0925/98
EARLIER FILING DATE: 1998-07-10
EARLIER APPLICATION NUMBER: 60/070,746
EARLIER FILING DATE: 1998-01-08
EARLIER APPLICATION NUMBER: 60/094,344
EARLIER FILING DATE: 1998-07-28
EARLIER APPLICATION NUMBER: 08/979,673
EARLIER FILING DATE: 1997-11-26
EARLIER APPLICATION NUMBER: 09/107,657
EARLIER FILING DATE: 1998-06-30
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 35
TYPE: PRT
ORGANISM: Talaromyces emersonii
US-09-199-290-4
```

```
Query Match 5.8%; Score 160; DB 3; Length 35;
Best Local Similarity 80.0%; Pred. No. 3,4e-07;
Matches 28; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 269 DANLLGSIHTPPEAACDSTFQPCSPRALANHK 303
D 1 DVNSILGSIHTPDPAGCDSTFQPCSPRALANHK 35
RESULT 22
US-09-821-616-4
Sequence 4, Application US/09821616
Patent No. 6620924
GENERAL INFORMATION:
APPLICANT: Nielsen, Bjarne R.
APPLICANT: Nielsen, Rudy
TITLE OF INVENTION: Thermostable Glucoamylase
FILE REFERENCE: 5279,200-US
CURRENT APPLICATION NUMBER: US/09/821,616
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/199,290
PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-24
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 1557/97
PRIOR FILING DATE: EARLIER FILING DATE: 1997-12-30
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 0925/98
PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-10
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/070,746
PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-08
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/094,344
PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-28
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/979,673
PRIOR FILING DATE: EARLIER FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/107,657
PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-30
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 35
TYPE: PRT
```

ORGANISM: Talaromyces emersonii  
US-09-821-616-4  
Query Match  
Best Local Similarity 5.8%; Score 160; DB 4; Length 35;  
Matches 28; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
Db 269 DANTLIGSIHTPDPEACDSTFPCSPRALANHK 303  
1 DVNSLIGSIHTPDAGCDSTFPCSPRALANHK 35  
RESULT 23  
US-09-134-001C-4463  
Sequence 4463, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
PRIOR FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4463  
LENGTH: 2137  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4463  
Query Match  
Best Local Similarity 5.6%; Score 156.5; DB 4; Length 2137;  
Matches 126; Conservative 92; Mismatches 284; Indels 79; Gaps 17;  
Db 2 SFRSLALGLVCTGLANYSKRATLDSML-----SNATVAKRAILNINIGADGAW 54  
1368 SEASATSLSGSTSTSLSDSTS-TSTSDASSTGTSVSDSNASTSLSGSLSTSVSDSTS 1426  
55 GADSGIVASPTDNDPYFTWTRDGLVKTLDVDFRNGDTLLST---INNYISAQA 110  
1427 TSS-----ASATSEBDSERASTSLSGSTSTSLSDSTSSTSDASTSVSSNSTS 1482  
111 IVQGISNPGDLSGAGLGPKNVDEATAYTSGMGRPORDPALRATAMIGCOMLIDNG 170  
1483 ISRLSTSVSDSTS-----TSTSDAST-STSVSDSASTSSSVSTS---DSE 1529  
171 YNSTADIVWPLVRNLSYVAQYNOTGYDLMEEVNGSFFTAVQHRALVEGSAFATV 230  
1530 STSTSTS-----DSASTSVSSNSTSTSL---SGSTSVSDSTSTSTSDASASTS 1580  
231 GSSCSMCDGQAPFLLCYLGFMTGSRFLANFDSRSGKANTLLGSIHTPDPEACDDST 290  
1581 ESDSDASTSSSSSVSTSVSDSTSASTSEASTSVSDSNASTSL-SESTSTSLSDST 1639  
291 FOPCSPRALANHKVEVDSFRSIYTLNDGLSD--SEAVAVGRYPEDTYNGNPFLLCTLA 348  
1640 SMSTSDASTSTSE-SDSDASTSLSDSTSVSESTSTSTSVSASNSTSTSLSDSR 1698  
349 AEGLYALVOMDQGLSEVTVLDFPKALYSDAATGYSSSST----- 393  
1699 TSLSDSTSTSTSESGSTSTSESDSDASTSLSESTSTSTSLSDSTSTSDASTSMVSDS 1758  
394 ---YSGIVADVKTFPAGFVSVTHAASNG-----SMSEQDKSDGQLSARDL 439  
1759 NRASSTSLSDSTSVSDSTSTSTSESTSTSTSESTSTSTSESTSTSVSDSTSTST 1816  
440 TWSYALLLTANNRNSVVPASMGETSASVPGTCAATSA-IGTVSSVTVTWPSIVATG 498  
1817 TSDSASTSTSESDNSTSTSLSESTSTSVSDSTSTSTSTSTSTSTSTSTSTSTSLSG 1876

499 TTT-----TATPTGSGSVTSTKTTATASKTSTTRSGML 534  
1877 TSTSVSDST 1917  
RESULT 24  
US-09-463-712C-10  
Sequence 10, Application US/09463712C  
Patent No. 6558937  
GENERAL INFORMATION:  
APPLICANT: DSM, N.V.  
APPLICANT: Gielkens, Marcus  
APPLICANT: Vessier, Jacob  
APPLICANT: De Graeff, Jeandert  
TITLE OF INVENTION: CELLULOSE DEGRADING ENZYMES OF  
FILE REFERENCE: 24615-20135.00  
CURRENT APPLICATION NUMBER: US/09/463,712C  
CURRENT FILING DATE: 2000-04-04  
PRIOR APPLICATION NUMBER: PCT/EP98/05047  
PRIOR FILING DATE: 1998-07-31  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10  
LENGTH: 536  
TYPE: PRT  
ORGANISM: Aspergillus niger  
US-09-463-712C-10  
Query Match  
Best Local Similarity 5.1%; Score 142; DB 4; Length 536;  
Matches 89; Conservative 48; Mismatches 136; Indels 120; Gaps 20;  
Db 205 VNGSFFTTAVOHRALVEGSAFATVAVSSCSMCDGQAPFLLCYLGFMTGSRFLANFDS- 263  
161 LMGALFYVAMDADGCTSEYSGNKAKEYGTGYCDSCCPDLKFRNG-----EANCDCG 213  
264 SRSQGANLTLG-----SIHTPDPEACDDSTFPCSPRALANHKVEVDSFRSIYTLN 316  
214 EBSNNVNVNGVDHSGCAEMDVMEANSISNAFTAHPC-----DSVQTMCDG 261  
317 DGLSDEAVAVGRY-----PEDTYN-----GN-----PMFICTLAAEQLY 353  
262 DSCGTYGYSAGRYSCPCPDGDIYPRLGNTDFRPGIYVDNPFVTVTOFTD--- 318  
354 DALYOMDKQGLEVTVDLSLDFPKALYS-----AATG--TVSS--SSSTYSSIYDAVATPA 405  
319 -----DGTSSGTLTEI-----KRLYVQNGEVIANGASTYSSVNGSITSAPCESKTLF 367  
406 DGFVSIIVETHAASNGSMSEQYDKSDGEQLS-----ARDLTWYALLLTANNRNSVVPAS 460  
368 -GDENVFDKGLLEG--MGEAMAKGVLVLSDVDYAADMLM-----LDSB 410  
461 WGBTASVSP-----GTCAATSAIG-----TVSSVTVTWPSIVANGGTTTAT 504  
411 YPVNSASTPGVARGICSDSGVPATVEASPNAYTVISNT-----KCPGISSTSSGS 464  
505 PTGSGSVTS-----TSKTTATASKTSTTRSGMS 533  
465 SSGSGSSSSSSSTTKATSTTLKTTSTTSSGS 497  
RESULT 25  
US-08-614-377A-7  
Sequence 7, Application US/08614377A  
Patent No. 5976864  
GENERAL INFORMATION:  
APPLICANT: Smith, John  
APPLICANT: Bingle, Wade H.  
APPLICANT: No. 5976864ellini, John F.  
TITLE OF INVENTION: EXPRESSION AND SECRETION OF  
TITLE OF INVENTION: HETEROLOGOUS

TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson PC  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/614,377A  
FILING DATE: 12-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/194,290  
FILING DATE: 09-FEB-1994  
CLASSIFICATION: 435  
APPLICATION NUMBER: US 07/895,367  
FILING DATE: 09-JUNE-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Tsao, Y. Rocky  
REGISTRATION NUMBER: 34053  
REFERENCE/DOCKET NUMBER: 08106/002001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1026 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-614-377A-7

Query Match 4.9%; Score 136; DB 2; Length 1026;  
Best Local Similarity 23.2%; Pred. No. 0.009;  
Matches 123; Conservative 73; Mismatches 225; Indels 110; Gaps 23;

26 TLDSWLSNEATVATATLNNIGADGAMVSGADSGIVASPTDNPDPFYTWTRDSGLVVK 85  
28 TLDAAY-----ATQQTGGLSDAALTLTKLVNSTTAAVATQTYQ----PFTGVAPSAAGLD 79  
86 TLVDLFRNGDTSLSLTIENTYISAQAIYQGISNPSGDLSSGAGLGEPRFNVDETAATYGSWG 145  
80 FLVDSTTN--TNDLN--DAYYSKFAQENRFINFSINLATAGAGATAF---AAAYT----- 128  
146 RPRDGPALATAMIGGQWLLDNGYSTATDVIWPLVRNDLSVYQYRNQGYD-LMEE 204  
129 -----GVSYAQVATAYDKIIIGNAVATAAGVDVA-----AAVAFLSRQANIDYLTAF 175  
205 VNGSSEFTIAVQHRALVEGSFAFATAVGSSCWCDSQAPELLCYLOSFWTGSFIANDSS 264  
176 VRANTPTTAADIDLAVKALIGTILNAA-----TVSGIGGYATATAAMIN-DLS 224  
265 RSGKDANTLLG-SHTFDPPEAACDSTFPQCSPRALANHKEVDSFRSIYTLNDGLDSE 323  
225 DGLSTNDNAGAVNLFTAYPSSGVSGSTL-----SLTGTDTTLT--- 262  
324 AVAVGRPEPTYNGNFWFLCTLAABQLYDALYQMDKQSLLEVTDVSLDFKALYSDA- 382  
263 -----GTANNDTFVAGEVAGATTLVGGDTLSGG-----AGTDV-LMWVQAAVATL 307  
383 ATGYSSSSSTYSIYDAVKTF--ADGFVSIIVETHAASNGSMSEQYDKSDGEQLSADLT 440  
308 PTGVITISGIETMNTSGAATILNTSSGVTGLTALNTTSGA-AQTVTAGAGQNLTA--TT 364

441 WSYAALLTANRRNSVVPASWGETSASSVPGTCAATSAIGTYS----SVTWSMPSIVAT 496  
365 AAQANNVAVDGRANVTASVTSGTTTGG--ANSAGTVSVSANSSTTTTGALIAVT 422  
497 GGTTTATPTGSGSVTST-----SKTA-----TASKTSTTTSG 531  
423 GGTAVTVTAQTAGNAVNTTLTQADVTVTGNSTTAAVTVTOTPAATAGATVAG 473

RESULT 26  
US-09-142-648B-7  
Sequence 7, Application US/09142648B  
Patent No. 6210948  
GENERAL INFORMATION:  
APPLICANT: Smt, John  
APPLICANT: Bingle, Wade H.  
APPLICANT: No. 6210948elini, John F.  
TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULOE  
FILE REFERENCE: 08106/002002  
CURRENT APPLICATION NUMBER: US/09/142,648B  
CURRENT FILING DATE: 1999-03-30  
PRIOR APPLICATION NUMBER: PCT/CA97/00167  
PRIOR FILING DATE: 1997-03-10  
PRIOR APPLICATION NUMBER: US 07/614,377  
PRIOR FILING DATE: 1996-03-12  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 1026  
TYPE: PRT  
ORGANISM: Caulobacter crescentus  
US-09-142-648B-7

Query Match 4.9%; Score 136; DB 3; Length 1026;  
Best Local Similarity 23.2%; Pred. No. 0.009;  
Matches 123; Conservative 73; Mismatches 225; Indels 110; Gaps 23;

26 TLDSWLSNEATVATATLNNIGADGAMVSGADSGIVASPTDNPDPFYTWTRDSGLVVK 85  
28 TLDAAY-----ATQQTGGLSDAALTLTKLVNSTTAAVATQTYQ----PFTGVAPSAAGLD 79  
86 TLVDLFRNGDTSLSLTIENTYISAQAIYQGISNPSGDLSSGAGLGEPRFNVDETAATYGSWG 145  
80 FLVDSTTN--TNDLN--DAYYSKFAQENRFINFSINLATAGAGATAF---AAAYT----- 128  
146 RPRDGPALATAMIGGQWLLDNGYSTATDVIWPLVRNDLSVYQYRNQGYD-LMEE 204  
129 -----GVSYAQVATAYDKIIIGNAVATAAGVDVA-----AAVAFLSRQANIDYLTAF 175  
205 VNGSSEFTIAVQHRALVEGSFAFATAVGSSCWCDSQAPELLCYLOSFWTGSFIANDSS 264  
176 VRANTPTTAADIDLAVKALIGTILNAA-----TVSGIGGYATATAAMIN-DLS 224  
265 RSGKDANTLLG-SHTFDPPEAACDSTFPQCSPRALANHKEVDSFRSIYTLNDGLDSE 323  
225 DGLSTNDNAGAVNLFTAYPSSGVSGSTL-----SLTGTDTTLT--- 262  
324 AVAVGRPEPTYNGNFWFLCTLAABQLYDALYQMDKQSLLEVTDVSLDFKALYSDA- 382  
263 -----GTANNDTFVAGEVAGATTLVGGDTLSGG-----AGTDV-LMWVQAAVATL 307  
383 ATGYSSSSSTYSIYDAVKTF--ADGFVSIIVETHAASNGSMSEQYDKSDGEQLSADLT 440  
308 PTGVITISGIETMNTSGAATILNTSSGVTGLTALNTTSGA-AQTVTAGAGQNLTA--TT 364  
441 WSYAALLTANRRNSVVPASWGETSASSVPGTCAATSAIGTYS----SVTWSMPSIVAT 496  
365 AAQANNVAVDGRANVTASVTSGTTTGG--ANSAGTVSVSANSSTTTTGALIAVT 422  
497 GGTTTATPTGSGSVTST-----SKTA-----TASKTSTTTREG 531  
423 GGTAVTVTAQTAGNAVNTTLTQADVTVTGNSTTAAVTVTOTPAATAGATVAG 473

RESULT 27  
US-08-728-470-10  
Sequence 10, Application US/08728470  
Patent No. 5928651  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Shoemaker and Matzare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,470  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstreiser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-633  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ. ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1529 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-728-470-10  
Query Match 4.8%; Score 133; DB 2; Length 1529;  
Best Local Similarity 20.3%; Pred. No. 0.031;  
Matches 119; Conservative 73; Mismatches 153; Indels 242; Gaps 28;  
QY 15 TGLANVISK-RATL--DSWLSNEATVARTALINNNIGADG-AWMSGADSGIV-----VASPS 66  
DB 777 TGNINITNKANVTLQADTNSNGLKKRITLTGNISVEGLSLTGNANANVGNLSIAEDS 836  
QY 67 T-----DNDPVEYVTRDGLVLTVDLFRNGDTSLSTIENYISAQAIYVGISNPSG 120  
DB 837 TTFGEASDNLNITGFT-----NNGTANI-----NIKGVYKLGQ 871  
QY 121 DLSSGAGLGEPRKNNDETATYTSWGRPRQDDPALRATAMIGFQMLDNGYSTATDIYW 180  
DB 872 DINNKGGL-----NI-----TTNASGTOKITINGITNEKGDINI 906  
QY 181 PLVARNLSVYAOYMNQGVYL-----WEVNGSSFTTAVQHRALVESAATAV 230  
DB 907 KNIKADAET-----QIGNISQKEGNTLSSPKVNTNQITL-----KAGVEG----- 949  
QY 231 GSSCSWCDSCAPILCYLQSFMTGSLANFDSRSQKDNATLGLSIHTFDPRAACDST 290

DB 950 GRS-----DSSEA-ENANL---TIQTEKLKAGD--- 974  
QY 291 FQPCSPRALANKEVVDSPRSITYTLN-DGLSDSEAVAVRYPEDPYNPNPFLCTLAIA 349  
DB 975 -----LNISGFNFAELTA----- 987  
QY 350 EQLYDALYQMDKQSLAEVTVDSLDFEKKALYSDAATGYSSSSSYSDAVATFADGF- 408  
DB 988 -----KNGS-----DLTI-----GNASGGMADAKKVTYPPKVDG-KISTDGHN 1024  
QY 409 VSIVETHAASNGSMSEQYDKSDGEOLISAPDLTWS-----YALLTANRRNSV 457  
DB 1025 VILNISEVKTNSGSSNAGNDNSTGLITISAKDVTYNNVNTSHKTIINISAAAGNVTTKEGTTI 1084  
QY 458 PASWGETSASVPGTCATSAIGTYSVTVTSMPSTV-----ATGCTTTATPTG-- 507  
DB 1085 NATTGSVEVTAQNGITKGNL-----TSQNVTVTAETENLVTTENAVINATSGVINISTKTGDI 1141  
QY 508 SCSVTSTG-----KTTATVASTSTYTTTRSGMSL 534  
DB 1142 KGIISTSGNVNITASGNTLKXSNITGQDVTVTAADGALTTTASGTI 1188

RESULT 28  
US-08-719-641-10  
Sequence 10, Application US/08719641  
Patent No. 6218141  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Shoemaker and Matzare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/719,641  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 16-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstreiser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-625  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ. ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1529 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-719-641-10



RESULT 30  
US-08-325-267A-2  
; Sequence 2, Application US/08325267A  
; Patent No. 5585271  
; GENERAL INFORMATION:  
; APPLICANT: WATARI, JUNJI  
; APPLICANT: TAKATA, YOSHIHIRO  
; APPLICANT: OGAMA, MASAHITO  
; APPLICANT: PENITILA, MERJA  
; APPLICANT: ONNELA, MAIJA-LEENA  
; APPLICANT: KERANEN, SIRKKA  
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST  
; TITLE OF INVENTION: CONTAINING THEM  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT  
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/325,267A  
; FILING DATE: 18-NOV-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP PCT/JP94/00290  
; FILING DATE: 24-FEB-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 38871/1993  
; FILING DATE: 26-FEB-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ. ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1537 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-325-267A-2

Query Match 4.7%; Score 130.5; DB 1; Length 1537;  
Best Local Similarity 20.0%; Pred. No. 0.051;  
Matches 119; Conservative 97; Mismatches 236; Indels 143; Gaps 27;

QY 11 GLVCTGLAVISKRAILDSWLSNEATYARPAIINIGADCAWYSGADSGIVASPESTDNP 70  
DB 769 GLIST-----TTEPW-TGTFSTSTSTEMTIVGTNGQPTD--ETIVIRIPTSSE- 814  
QY 71 DYFYTWTRD--SGLVAKTIVDLFRNGDTSLSTIENYI-----SAQIVOGISNP-SGDL 122  
DB 815 --LVTTTTEMTTFTSTSTEMTITGTNGVPTDEVIIVIRTPSGLISTTTEMTTGT 872  
QY 123 SSGAGLGEPRKENVDETAATYGSWGRPORD-----GPALRTAMIGFGQWLLDNGYSTAT 176  
DB 873 TS-----TSTEMTITGTNGQPTDEVIIVIRTPSGLISTTTEPW--TGTFSTST 922  
QY 177 D-----IVPLVANDLSYVAQYNNQGYDLMEV-----NGS-SFP 211  
DB 923 EMTTIVGTNGVPTDEVIIVIRTPSGLISTTTEPWGTFTSTSTETVITITGTNGQPTDE 982  
QY 212 TIAVQRALVEGSAFATAVAGSSCWCDQAPILICVLOSFWTOSFILANPD-----SSRSG 267

DB 983 TVIVIRTPSGLISTTT-----||||| : : : :  
QY 268 KOANTLIGSIHFPDEPAACDDSTFOCPSPALANKHEVDSRSRYTLNDGLSDSPAVV 327  
DB 1023 OPTDETVIVIRTPSGLVTTT--EPMGTFTSTSTEM-----STVGTNGLPPTDETVIV 1076  
QY 328 GRYP-----EDTYVYNGNPMWFLCTLAABQYDALYQMDKQSGLEVTVDELDFKALYSDA 382  
DB 1077 VKTPTTAISSSSSSSGQITSISRSPIITPFYNSGTSVSISSSVTSSLSFTSS 1136  
QY 383 ---ANGTSSSSSTYSIV-DAVKTFADGFVSIIVETHAASNGSMSEQYKSDG----- 431  
DB 1137 PVISSSVISSSTTSTISPSKSS-----SVIFPSSSTSGS-SESETSAGSVSSSR 1189  
QY 432 -EOLSRDLTWSYALALTANRNRSVVPASWGETSASVPGCAATSAIGTSVTVTSW 490  
DB 1190 ISSESSKSPYSSSSLPVTSATTS-----OETASLPAPATTTXTSEQTTL--VTVBC 1241  
QY 491 PSIVATGTT---TTATPTSGSVT-----STSKTATASKTSTTR 529  
DB 1242 ESHVCTESISPAIVSIAIVTSGVTTETTWCPISITTEITTKOTKGTTEOTTEITTK 1296

RESULT 31  
US-08-617-697-10  
; Sequence 10, Application US/08617697  
; Patent No. 5977336  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Maltare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/617,697  
; FILING DATE: 01-APR-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 05-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Berklesseer, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-557  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ. ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1600 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-617-697-10

Query Match 4.7%; Score 130; DB 2; Length 1600;  
Best Local Similarity 20.4%; Pred. No. 0.06;











```

Db      490 SNLFQSSGR-----SPWNSINFIIDVHDGMTLKDVSCH--GANNQAMPYG--PSDG 537
Qy      335 YVNGNPFMLCTLLAAEQLYDALYQWDKQSLVETDVSLDFEKLAYSDA-----TGT--- 386
Db      538 GISTN-----YSWD-QGMSAGTGAIVDQRRAARGMAFEMLSAGTPLM 579
Qy      387 -----YSSSSS-----TYSSIVDAVK--TFADGFVSIIVETHAANGSMSE 424
Db      580 QGDEYLRTLCNNNAYNLSDSANMLTYSWTTDQSNFYFAQRLIAFRKAHPLRPS--- 636
Qy      425 QYDKSDGEQLSARDLTWMSYAAALLTANNRRNSVVPASGETS---ASSVGTCAATSAG 480
Db      637 -----SWYSGSQLTWQPSGAVADSN-----YMNNTSNVAIAIYAING-----PSLG 677
Qy      481 TYSSVTV--TSWPSIYATGTTTATPTGSG--SVTST-----SKTATASKT----- 524
Db      678 DSNISIVAYNGWSSSVT---FTLPAPPSGTQWYRVTDTCMDWDGASTFVAPGSETLIGA 734
Qy      525 -STTTRSGMSL 534
Db      735 GTTYGCGQGS 745

```

## RESULT 38

```

US-09-346-237-7
; Sequence 7, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Biagard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629-200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; EARLIER FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 776
; TYPE: PR1
; ORGANISM: Pseudomonas species SMP1
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(776)
; OTHER INFORMATION: Isoamylase
US-09-346-237-7

```

```

Query Match      4.6%; Score 127.5; DB 3; Length 776;
Best Local Similarity 21.8%; Pred. No. 0.031;
Matches 146; Conservative 67; Mismatches 213; Indels 245; Gaps 38;

```

```

Qy      45 NIGADGAWYSGADSGI-----VVASPTNDPDVFFYMTDRSGVLTIVLDFNGSTSL 99
Db      164 NVFASASASTRTDSTGIYAPKGVLPVSTOSTGKTPRAQKDVYEHVHVGFEQDTSIP 223
Qy      100 STE-----NYISAQAI-----VQGISNPSGDL--SSGAG-----LGEPEKN 134
Db      224 AQVRYGTYGAGLASYLASGYTAVFELPQGEONANDVPRSDANQNWGWTENYS 283
Qy      135 VDET-AYTSWGEPQDGPALRATAMI-----GF-----GQMLDNGYTS 173
Db      284 PRRRYVYNKAAG-----GPTAEFOAWVOAFHNAGIKVYMDVYVNHTRAEGETWSSD--PT 336
Qy      174 TATDIWPLVRNDLSVVAQWNTQGYDIMEEVNGSSFFTI-AVQHRALVEGSAF----- 226
Db      337 TATTYWREGIDNTTYIELISGNQYFD--NTGIGANFNNTYVNAQMLIVDSGLAWYANTWG 394
Qy      227 -----ATVAVSSC--SWCDSQAP-----ILCYLOSF----- 251

```

```

Db      395 VDFRFDLASVLGNSCLNGAYTASAPNCPNGYNFDPADSNVAINRLIBEFFVRPAAGS 454
Qy      252 -----WT-----GSFILANF-----DSSRSG-----KDNATLLGS 276
Db      455 GDLFAEPALIGNSYQLGFFPGWSENNGLPRDSLRQONELGSMITIVYTOANDPSSG 514
Qy      277 IHTEDEAACDSTFQPCSPRALANKEVVD--SFRSIYTLNDGLDSEAVAVGRYPEDT 334
Db      515 SNLFQSSGR-----SPWNSINFIIDVHDGMTLKDVSCH--GANNQAMPYG--PSDG 562
Qy      335 YVNGNPFMLCTLLAAEQLYDALYQWDKQSLVETDVSLDFEKLAYSDA-----TGT--- 386
Db      563 GISTN-----YSWD-QGMSAGTGAIVDQRRAARGMAFEMLSAGTPLM 604
Qy      387 -----YSSSSS-----TYSSIVDAVK--TFADGFVSIIVETHAANGSMSE 424
Db      605 QGDEYLRTLCNNNAYNLSDSANMLTYSWTTDQSNFYFAQRLIAFRKAHPLRPS--- 661
Qy      425 QYDKSDGEQLSARDLTWMSYAAALLTANNRRNSVVPASGETS---ASSVGTCAATSAG 480
Db      662 -----SWYSGSQLTWQPSGAVADSN-----YMNNTSNVAIAIYAING-----PSLG 702
Qy      481 TYSSVTV--TSWPSIYATGTTTATPTGSG--SVTST-----SKTATASKT----- 524
Db      703 DSNISIVAYNGWSSSVT---FTLPAPPSGTQWYRVTDTCMDWDGASTFVAPGSETLIGA 759
Qy      525 -STTTRSGMSL 534
Db      760 GTTYGCGQGS 770

```

## RESULT 39

```

US-08-797-366-3
; Sequence 3, Application US/08797366
; Patent No. 5853702
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Christgau, Stephan
; APPLICANT: Halkier, Torben
; APPLICANT: Shuster, Jeff
; APPLICANT: Fuglsang, Claus Crone
; TITLE OF INVENTION: Penicillium Purpurogenum Mutanases
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSER: No. 58537020 No. 5853702disk of No. 5853702ch America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; APPLICATION NUMBER: US/08/797,366
; FILING DATE: 09-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-878-9655
; TELEFAX: 212-867-0123
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 630 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-797-366-3

```

Query Match 4.6%; Score 127; DB 2; Length 630;  
Best Local Similarity 21.5%; Pred. No. 0.025;  
Matches 131; Conservative 77; Mismatches 229; Indels 172; Gaps 32;

```
QY 2 SFRSLALSLGVTCTGLA--NVISKRAITLDSMLSEAVIA----RTAILNNIGAD--GAMV 53
DB 6 AFATATLAIILAACSLALPSDSMVSRRSTSDRLVFHFVNGVVSRTS--ASDYADMQAKA 64
QY 54 SGADS-GIVVASPSTNDPFIYTWTRDGLVLTVDLFRNGDTSLSTIENTISAQAIY 112
DB 65 YGIDAFALNIGTDFPSQOQAGYAESANNMKVFISF----DFNMWSTGQATEIGKIA 120
QY 113 QGISNPSGDLSSGAGLCEPKFENVDEIAYTGSWGRPQRDPALRATAMIGFGWMLDNGYT 172
DB 121 QYGSLLPG-----QIMYDDKIFVSSFAGDGVVALKSA--GGNVFAPAFPH 166
QY 173 STATDIY-----WPLVRNDLSYVAQYMNQGYDLMEEVNGSSFPIAVQHRALVE 222
DB 167 SYGTDLSDVDGLLNMGMFSGNKKAPTA-----GANVVEEGDEEYIT-ALDGKPYIA 219
QY 223 GSA--FATAVSSCSCWCSQA-PEILCYLOSFWTGSFILANF-----DSSR 265
DB 220 PASFWFSTHFGPBTYTSKNWVFPSDLFFYQR-WND--LNLGPOFIEVVTWMDYGSQY 275
QY 266 SGKDANTLLGSIHTFDEPA--ACD--DSTFQPCSPALANHKRVDSFRSIYTLNDGL- 319
DB 276 VGP-----LNSPHTDSSSRWADMPHDGWLDAKPYIAAFH-DGATSLSSSYTTEDQLI 329
QY 320 -----SDSEAVAVGRYPEDT--YVNGNP--WFLCTLAALAEQLYDALYOWDK 361
DB 330 YWRPQRLMDCDATTICWVAAN--NDTGNVFEGRNGW-----ESMEDAV----- 373
QY 362 QGSLVETVDSLDFFKALYSDAATGYSSSSSTYSIVDAVKTADGFSIVETHAANGS 421
DB 374 -----FVALLQSGAGTYQVTSQPT-----ETDAPAGA 402
QY 422 MSEQYD-----KSDEQL-----SARDLT-----WSYALLTANNRRNSV 457
DB 403 SAFQVPMGFQPSFSLSRDGETVLGSLKDIIDGCLGYNFNAVYGSJPA-TFSDPLE 461
QY 458 PASGERSASVPTCAATSAIGYSSVTVSWPSIYATGTTTATPTGSGVTSYST 517
DB 462 PPSLNAPSEGLKSTCATPSLG---LTSTTPEETIPGTITGSAITGAATTSTIST 517
QY 518 TATASKTST 526
DB 518 TSTISTST 526
```

RESULT 40  
US-08-956-268-3  
Sequence 3, Application US/08956268  
Patent No. 5874275  
GENERAL INFORMATION:  
APPLICANT: Berka, Randy M.  
APPLICANT: Christgau, Stephan  
APPLICANT: Halkier, Torben  
APPLICANT: Shuster, Jeff  
APPLICANT: Fuglsang, Claus Crone  
TITLE OF INVENTION: Penicillium Purpureogenum Mutasees  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5874275 No. 5874275disk of No. 5874275th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/956,268  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/797,366  
FILING DATE: 09-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4593.200-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 630 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-956-268-3

Query Match 4.6%; Score 127; DB 2; Length 630;  
Best Local Similarity 21.5%; Pred. No. 0.025;  
Matches 131; Conservative 77; Mismatches 229; Indels 172; Gaps 32;

```
QY 2 SFRSLALSLGVTCTGLA--NVISKRAITLDSMLSEAVIA----RTAILNNIGAD--GAMV 53
DB 6 AFATATLAIILAACSLALPSDSMVSRRSTSDRLVFHFVNGVVSRTS--ASDYADMQAKA 64
QY 54 SGADS-GIVVASPSTNDPFIYTWTRDGLVLTVDLFRNGDTSLSTIENTISAQAIY 112
DB 65 YGIDAFALNIGTDFPSQOQAGYAESANNMKVFISF----DFNMWSTGQATEIGKIA 120
QY 113 QGISNPSGDLSSGAGLCEPKFENVDEIAYTGSWGRPQRDPALRATAMIGFGWMLDNGYT 172
DB 121 QYGSLLPG-----QIMYDDKIFVSSFAGDGVVALKSA--GGNVFAPAFPH 166
QY 173 STATDIY-----WPLVRNDLSYVAQYMNQGYDLMEEVNGSSFPIAVQHRALVE 222
DB 167 SYGTDLSDVDGLLNMGMFSGNKKAPTA-----GANVVEEGDEEYIT-ALDGKPYIA 219
QY 223 GSA--FATAVSSCSCWCSQA-PEILCYLOSFWTGSFILANF-----DSSR 265
DB 220 PASFWFSTHFGPBTYTSKNWVFPSDLFFYQR-WND--LNLGPOFIEVVTWMDYGSQY 275
QY 266 SGKDANTLLGSIHTFDEPA--ACD--DSTFQPCSPALANHKRVDSFRSIYTLNDGL- 319
DB 276 VGP-----LNSPHTDSSSRWADMPHDGWLDAKPYIAAFH-DGATSLSSSYTTEDQLI 329
QY 320 -----SDSEAVAVGRYPEDT--YVNGNP--WFLCTLAALAEQLYDALYOWDK 361
DB 330 YWRPQRLMDCDATTICWVAAN--NDTGNVFEGRNGW-----ESMEDAV----- 373
QY 362 QGSLVETVDSLDFFKALYSDAATGYSSSSSTYSIVDAVKTADGFSIVETHAANGS 421
DB 374 -----FVALLQSGAGTYQVTSQPT-----ETDAPAGA 402
QY 422 MSEQYD-----KSDEQL-----SARDLT-----WSYALLTANNRRNSV 457
DB 403 SAFQVPMGFQPSFSLSRDGETVLGSLKDIIDGCLGYNFNAVYGSJPA-TFSDPLE 461
QY 458 PASGERSASVPTCAATSAIGYSSVTVSWPSIYATGTTTATPTGSGVTSYST 517
DB 462 PPSLNAPSEGLKSTCATPSLG---LTSTTPEETIPGTITGSAITGAATTSTIST 517
QY 518 TATASKTST 526
DB 518 TSTISTST 526
```

RESULT 41

US-09-717-364A-15  
; Sequence 15, Application US/09717364A  
; Patent No. 6663872  
; GENERAL INFORMATION:  
; APPLICANT: Pitkovski, Jacob  
; APPLICANT: Muelem, Margalit  
; APPLICANT: Koren, Ziv Reu  
; APPLICANT: Krippe, Simcha  
; APPLICANT: Shmueli, Esther  
; APPLICANT: Perez, Yifat  
; APPLICANT: Guter, Bezalel  
; APPLICANT: Gaili, Galid  
; APPLICANT: Michael, Amnon  
; APPLICANT: Goldberg, Doron  
; TITLE OF INVENTION: HEMORRHAGIC ENTERITIS VIRUS DNA SEQUENCES, PROTEINS ENCODED THERE  
; TITLE OF INVENTION: VARIOUS USES THEREOF  
; FILE REFERENCE: 1567/6365  
; CURRENT APPLICATION NUMBER: US/09/717,364A  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: IL124567  
; PRIOR FILING DATE: 1998-05-20  
; PRIOR APPLICATION NUMBER: PCT/IL9900268  
; PRIOR FILING DATE: 1999-05-19  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: Patent version 3.1  
; SEQ ID NO 15  
; LENGTH: 906  
; TYPE: prf  
; ORGANISM: hemorhagic enteritis virus  
US-09-717-364A-15

Query Match 4.6%; Score 127; DB 4; Length 906;  
Best Local Similarity 21.8%; Pred. No. 0.044;  
Matches 97; Conservative 58; Mismatches 155; Indels 134; Gaps 21;

QY 155 RATAMGCOMLNDNGYSTAIDVWPLVR-----NDLSYVAQYNNQSYDLM 202  
DB 83 RTIAGVGR-VLDMG--STYFDIRGNIDRGPFKYGAVNPLAPRGAQFNMI----- 134  
QY 203 EGVNSSFPTIAGVAGALVEGSAFATAVSSGSCWCSQAPELLCYQSTWTSFILAND 262  
DB 135 KTVGGKTYLT-----AQATKFFSTSGNGCAANTEASSFTNLVPSPTGS----- 179  
QY 263 SRSRGKANTLLGSIHTFDPBEAACDSTFOPCSRPLAHKVVDSFRSITLNDGLSDS 322  
DB 180 -----AESFDP-----TEGASCRAT-----TL--GSSVT 203  
QY 323 EAVAVGRYPEDTYNGN---PWFLCTIAAEQLYDALYOMDKOSLEVT-DVSLDFKAL 378  
DB 204 DATCGAVPIPIQANGSILP-----SVTPDKKPAD-----GKSGSVTCTRAICCDNVTVQ 255  
QY 379 YSDAATGYSSSSSTYSIVDAVK--TFADGVYSIV-ETHAASNGM-SEQYDKSGEQL 434  
DB 256 YPDRIVAYDSYDKLATRMGNRINYIGFRDNFGLMYNNGAHSGLATETGDIINVEQL 315  
QY 435 SARDLTWSYALLTANRRNSVVPASWGET-----SASSVPGCATSAI 479  
DB 316 QRNRTEISIQWMLADMSRHHY--SQMNOAVDDYDLNRYLTNIGYEEGPGCYPTGCM 374  
QY 480 GTYSSVTVTSPSIATGTTTATPT-----GSGSVTST----- 514  
DB 375 GNYPN-TWMSVGLVNNGTATATTNTVAVMGFSVPTMEINVOAYLQCKMYANIAEY 433  
QY 515 -----SKTATATASKSTTTRSGMS 533  
DB 434 LPDKYKKAIGTSEIDPTTYSYMN 457

RESULT 42  
US-09-346-237-4  
; Sequence 4, Application US/09346237A  
; Patent No. 6265197  
; GENERAL INFORMATION:

; APPLICANT: Biegard-Frantzen, Henrik  
; APPLICANT: Svendsen, Allan  
; TITLE OF INVENTION: Starch Debranching Enzymes  
; FILE REFERENCE: 5629.200-US  
; CURRENT APPLICATION NUMBER: US/09/346,237A  
; CURRENT FILING DATE: 1999-07-01  
; EARLIER APPLICATION NUMBER: PA 1998 00868  
; EARLIER FILING DATE: 1998-07-02  
; EARLIER APPLICATION NUMBER: 60/094,353  
; EARLIER FILING DATE: 1998-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for windows version 3.0  
; SEQ ID NO 4  
; LENGTH: 776  
; TYPE: prf  
; ORGANISM: Pseudomonas amylocleromosa  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)..(776)  
; OTHER INFORMATION: Isoamyase  
US-09-346-237-4

Query Match 4.6%; Score 126.5; DB 3; Length 776;  
Best Local Similarity 21.8%; Pred. No. 0.038;  
Matches 146; Conservative 67; Mismatches 213; Indels 245; Gaps 38;

QY 45 NIGADGAWISGADSGI-----VVASPSTNDPDIYFTWTRDSGLVLTLYDLFRNGDTSL 99  
DB 164 NVFASGASYRTTDSIGYAPKGVLVSTOSTKPRACKDDVIYEVHVGFTEDOTSIP 223  
QY 100 STIE-----NYSHQAI-----VQGISNPSGL--SSGAG-----LGEPRKN 134  
DB 224 AQYRGTYAGAKASTIASLGYTAVEFLPVQETQDANDVENSNDANQWKGMTENTYS 283  
QY 135 VDET-AYTGSGWRPORDGPALRATMI-----GF-----GQWLLDNGYS 173  
DB 284 PRRRAYNKAAG-----GPTAEQAMVQAFHNAQIKVYMDVYNNHTABEGTWTSSD--PT 336  
QY 174 TATDIWPLVRNDLSYVAQYNNQSYDLMEEVNGSSFTI-AVQHALVEGSAF----- 226  
DB 337 TATYISWRGLDAVYELTSGNQFYD--NTGIGANFNTYNTVAQNLIVDSLAWYMWNTWG 394  
QY 227 -----ATAVGSSC--SWCSQAPB-----ILCYLOS----- 251  
DB 395 VDGFRDLASVUGNSCLNCAATYASAPNCPCNGYNDADSNVAIRILREFTVRRPAGS 454  
QY 252 -----WT-----GSFTIANF-----DSSRSG-----KDANTLGS 276  
DB 455 GDLFAEPWALIGNSYQGGFQGWSEWNGLFPRDSLQANQNELGSWTIVIQDANDFGS 514  
QY 277 HTFDPBEAACDSTFOPCSRPLAHKVVYD--FRSITLNDGLSDSEAVAVGRYPEDT 334  
DB 515 SNLFQSSGR-----SPWNSINFIDVHDMTKLDVYSCN-GANNSSQAMPYG--PSDG 562  
QY 335 YNNGNPMFLCTIAAEQLYDALYOMDKOSLEVTQDVSLDFKALYSDDA-----TGT 386  
DB 563 GTSTN-----YSMD-QGMSAGTAANDQRAARTGMAFEMLSACTPLM 604  
QY 387 -----YSSSSS-----YSSSIYDAVK--TFADGVYSIVETHAASNGMSGE 424  
DB 605 QGDEYLRTLQCNNAVNDSSANWLTYSWTTDQSFYFAQRLAIFRAHAPALRS----- 661  
QY 425 QYDKSDGQLSARDUTWYSYALLTANRRNSVVPASWGETS-----ASSVPGCATSAIG 480  
DB 662 -----SWYSGSQLTWYQPSGAVADSN-----YMNNTSYAIAIYALNG-----PSLG 702  
QY 481 TYSSVTV--TSPSIATGTTTATPTGSG--SVTST-----SKTATATASKT----- 524  
DB 703 DSNSTIYAVNGSSSVT---FTLPAPPSGTQMYRYTDTCDMDAGASTFVAPSPSEITLIGA 759  
QY 525 -STTRSGMSL 534  
DB 760 GTTYGCGGSL 770



```

Qy 184 RNDLSVYAOYMNQGYDLWEVNGSSFFITAVQHRALVBSAPATAVSSC-SWCDSQAP 242
Db 141 -----TGFFLPQ--TGSYTFKAT-----VDSALISVGATAFNCAQOQP 181
Qy 243 EILCYIOF-----WTGS-----FILANFSSRSKCDANTLGSIHTEPDAAC 286
Db 182 PITS--TNEFTIDGIKPMGSLPENIEGTVMYAGYYPMKVYSNAV--SMGTLPISVTL 237
Qy 287 DSTFQPCSPBALANHKEVVDSPRS-IYTLNDGLSDSEAVAVGRYPEDTYY-----NG 338
Db 238 PGCT-----TVSDDEGYVSGFDDLSQSNCTV---PDPSNVAVSTTTT 280
Qy 339 NPFLLCTLAAEQLYDALYQMDKQSLLEVTVSLDFFKALYSDAATGYSSSSSTYSIV 398
Db 281 BPW-----TGTFTSTSTEMT 298
Qy 399 DAVKTFADGVSIIVETHAASNGSMSEQYDKSDGEQLSARDLTWSYALLTANNR--NS 455
Db 299 GNGVPTDETVIVIRT-PTSEGLISTTBEPWTGFTS---TSTEVTITGTNGQPTDET 353
Qy 456 VV-----PASMGETSASSVPCTCAATSAIGTSSVTVTSMPSIVATGTTT-----TATP 505
Db 354 VIVIRPTSEGLISTTBEPWT-----GFTS-TSTEMTGTGTNGQPTDETVIVIRP 405
Qy 506 TSGSVYTSKT-----TATASKTSTT 528
Db 406 TSEGLVTTTTEPWTGFTSTSTEMSTVT 433

```

Search completed: June 28, 2004, 07:45:58  
 Job time : 28 secs



**THIS PAGE LEFT BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 07:39:37 ; Search time 60 Seconds  
(without alignments)  
2514.673 Million cell updates/sec

Title: US-10-038-723-2

Perfect score: 2771

Sequence: 1 MSFSLALSGLVCTGLANV.....SKTTATASKSTYTRSGMSL 534

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database: A\_Geneseq\_290an04:\*

1: geneseqp1980s:\*\n2: geneseqp1990s:\*\n3: geneseqp2000s:\*\n4: geneseqp2001s:\*\n5: geneseqp2002s:\*\n6: geneseqp2003as:\*\n7: geneseqp2003bs:\*\n8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2771	100.0	534	2	AAV18090 Truncated
2	2771	100.0	534	2	AAV23338 Aspergill
3	2771	100.0	534	3	AAV03450 Aspergill
4	2771	100.0	534	3	AAV77740 A. niger
5	2771	100.0	534	4	AAV48171 A. niger
6	2771	100.0	534	4	AAV61904 A. niger
7	2742	99.0	640	4	AAV77741 A. niger
8	2742	99.0	640	4	AAV61905 A. niger
9	2683.5	96.8	639	1	AAV0212 Sequence
10	2658	95.9	630	1	AAV81876 Sequence
11	2658	94.8	616	3	AAV15176 Aspergill
12	2655	94.7	616	2	AAV55979 Aspergill
13	2655	94.7	616	3	AAV5180 Aspergill
14	2615	94.4	616	2	AAV55976 Aspergill
15	2615	94.4	616	3	AAV5178 Aspergill
16	2614	94.3	616	3	AAV15184 A. awamori
17	2612	94.3	616	2	AAV55980 Aspergill
18	2604	94.0	616	2	AAV15181 Aspergill
19	2604	94.0	616	2	AAV55977 Aspergill
20	2600.5	93.8	621	2	AAV55978 Aspergill
21	2589.5	93.5	621	3	AAV15183 A. awamori
22	2584.5	93.3	621	3	AAV15182 A. awamori
23	2523	91.1	1095	6	ABP96630 Alpha-amy
24	2468	89.1	471	4	ABM00045 AMG SEQ I

26	1850.5	66.8	631	6	ABE80181	AbB80181 A. fumiga
27	1770	63.9	624	4	AAV48170	AAV48170 Thermococ
28	1689	61.0	618	2	AAV23339	AAV23339 Talacomye
29	1674.5	60.4	591	2	AAV23337	AAV23337 Talacomye
30	1615.5	58.3	630	5	AAV51596	AAV51596 Thielavia
31	1577	56.9	626	2	AAV71034	AAV71034 N. crassa
32	1425.5	51.4	581	3	AAV18823	AAV18823 Amino act
33	1425.5	51.4	581	5	AAV79444	AAV79444 Fusarium
34	1425.5	51.4	581	7	ABO01925	ABO01925 Fusarium
35	1408.5	50.8	704	6	ABE80170	ABE80170 A. fumiga
36	1380.5	49.8	620	6	ABE80174	ABE80174 A. fumiga
37	1321	47.7	616	6	AAV30155	AAV30155 Glucoamyl
38	725	26.2	579	6	ABP96632	ABP96632 Rhizopus
39	724	26.1	604	1	AAV60359	AAV60359 Glucoamyl
40	641.5	23.2	624	2	AAV77674	AAV77674 Glucoamyl
41	615.5	22.2	497	1	AAV70572	AAV70572 Glucoamyl
42	600.5	21.7	918	1	AAV60723	AAV60723 Sequence
43	374	13.5	293	1	AAV70183	AAV70183 Sequence
44	181	6.5	1588	7	ADC01413	ADC01413 Enterohae
45	172	6.2	32	2	AAV22343	AAV22343 His(184)
46	170	6.1	32	2	AAV22341	AAV22341 Ala(182)
47	169	6.1	32	2	AAV22342	AAV22342 Lys(183)
48	169	6.1	32	2	AAV22340	AAV22340 His(183)
49	169	6.1	32	2	AAV37120	AAV37120 Staphyloc
50	164.5	5.9	2344	4	AAV22345	AAV22345 Lys(183)
51	164	5.9	32	2	AAV22346	AAV22346 Ala(182)
52	164	5.9	32	2	AAV22346	AAV22346 Ala(182)
53	164	5.9	2016	6	AAV36891	AAV36891 pleiotreux
54	163	5.9	32	2	AAV22339	AAV22339 Arg(178)
55	161	5.8	32	2	AAV22344	AAV22344 Ala(182)
56	161	5.8	1063	5	ABE54168	ABE54168 Lactococ
57	160	5.8	35	2	AAV23334	AAV23334 Peptide d
58	159.5	5.8	2261	6	ABJ18914	ABJ18914 Pathogen
59	159.5	5.8	2271	6	ABW72734	ABW72734 Staphyloc
60	159.5	5.8	2283	6	ABP56876	ABP56876 Staphyloc
61	159	5.7	2271	6	ABU16000	ABU16000 Protein e
62	158	5.7	2870	3	AAV95559	AAV95559 Caenorhab
63	158	5.7	3178	3	AAV95556	AAV95556 Caenorhab
64	156.5	5.6	2137	5	ABP39618	ABP39618 Staphyloc
65	155.5	5.6	596	6	ABE59339	ABE59339 Human sec
66	154.5	5.6	1031	4	AAV35339	AAV35339 Enterococ
67	152	5.5	45	6	ABP96634	ABP96634 Raw-starc
68	151.5	5.5	1099	6	ABU02096	ABU02096 Protein e
69	147.5	5.3	4440	6	ABU88256	ABU88256 Novel hum
70	147.5	5.3	4440	6	ABU90135	ABU90135 Novel hum
71	147.5	5.3	4440	6	ABU96437	ABU96437 Novel hum
72	147.5	5.3	4440	6	ABU99046	ABU99046 Novel hum
73	147.5	5.3	4440	6	ABU98261	ABU98261 Novel hum
74	147.5	5.3	4440	6	ABU91967	ABU91967 Novel hum
75	147.5	5.3	4440	6	ABU85271	ABU85271 Novel hum
76	147.5	5.3	4440	6	ABU80410	ABU80410 Novel hum
77	147.5	5.3	4440	6	ABU88961	ABU88961 Novel hum
78	147.5	5.3	4440	6	ABU06457	ABU06457 Novel hum
79	147.5	5.3	4440	6	ABU95517	ABU95517 Novel hum
80	147.5	5.3	4440	6	ABU95207	ABU95207 Novel hum
81	147.5	5.3	4440	6	ABU90755	ABU90755 Novel hum
82	147.5	5.3	4440	6	ABU93917	ABU93917 Novel hum
83	147.5	5.3	4440	6	ABU86191	ABU86191 Novel hum
84	147.5	5.3	4440	6	ABU82046	ABU82046 Novel hum
85	147.5	5.3	4440	6	ABU07907	ABU07907 Novel hum
86	147.5	5.3	4440	6	ABU94227	ABU94227 Novel hum
87	147.5	5.3	4440	6	ABO00100	ABO00100 Novel hum
88	147.5	5.3	4440	6	ABU87111	ABU87111 Novel hum
89	147.5	5.3	4440	6	ABU91352	ABU91352 Novel hum
90	147.5	5.3	4440	6	ABU90445	ABU90445 Novel hum
91	147.5	5.3	4440	6	ABU97036	ABU97036 Novel hum
92	147.5	5.3	4440	6	ABO05232	ABO05232 Novel hum
93	146.5	5.3	4469	6	ABU42529	ABU42529 Protein e
94	146.5	5.3	1439	5	ABU16643	ABU16643 Protein e
95	144.5	5.2	614	5	ABE54071	ABE54071 Lactococ
96	144.5	5.2	738	7	ADC01840	ADC01840 C. albica
97	143.5	5.2	596	4	AAV99408	AAV99408 Human PRO
98	143.5	5.2	596	4	AAV66157	AAV66157 Protein o

99 143.5 5.2 596 4 AAU29178  
100 143.5 5.2 596 4 AAB87575

Aau29178 Human PRO  
Aab87575 Human PRO

## ALIGNMENTS

## RESULT 1

AA18090 ID AA18090 standard; protein; 534 AA.

XX AA18090;

XX 06-AUG-1999 (first entry)

XX Truncated A. niger glucoamylase G1 protein sequence.

XX Glucoamylase; saccharification; liquefied starch solution; cyclodextrin;  
XX enzymatic saccharification; monosaccharide production; oligosaccharide;  
XX dextrose; trehalose; isomaltoligosaccharide; maltoligosaccharide.

XX Aspergillus niger.

XX MO9927124-A1.

XX 03-JUN-1999.

XX 23-NOV-1998; 98MO-US024871.

XX 26-NOV-1997; 97US-00979673.

XX 30-JUN-1998; 98US-00107657.

XX (NOVO ) NOVO-NORDISK AS.

XX (STAL ) STALEY MFG CO A E.

XX liaw GC, Pedersen S, Hendriksen HV, Svendsen A, Nielsen BR;

XX Nielsen RI;

XX WPI; 1999-357849/30.

XX N-PSDB; AAX76983.

XX Saccharification of liquefied starch solution for production of, e.g.

XX monosaccharides.

XX Claim 18; Page 55-58; 63pp; English.

XX This sequence is a truncated A. niger glucoamylase G1 protein. The  
XX invention relates to a method of saccharifying a liquefied starch  
XX solution comprising: (a) a saccharification step during which one or more  
XX enzymatic saccharification stages (using a mutated Aspergillus niger  
XX glucoamylase) take place; (b) one or more high temperature membrane  
XX separation steps; and (c) re-circulation of the saccharification enzyme;  
XX in which the membrane separation steps are carried out as an integral  
XX part of the saccharification step. The method can be used for the  
XX production of mono and/or oligosaccharides from starch, including  
XX dextrose, trehalose, isomaltoligosaccharides, cyclodextrins or  
XX maltoligosaccharides. The use of the membrane separation step improves  
XX the efficiency of the process and improves yields and purity

XX Sequence 534 AA.

XX Query Match 100.0%; Score 2771; DB 2; Length 534;

XX Best Local Similarity 100.0%; Pred. No. 1.5e-214;

XX Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 MSRRSLALSLGLVCTGLANVSKRATIDSWLSNEATYARFAIINMGADGAWVSGADSGI 60

XX 1 MSRRSLALSLGLVCTGLANVSKRATIDSWLSNEATYARFAIINMGADGAWVSGADSGI 60

XX 61 VVASPTDNPDYFYTWTRDSGLVLTVDLFRNGDTSLSLTENYISAQAVOGISNPSG 120

XX 61 VVASPTDNPDYFYTWTRDSGLVLTVDLFRNGDTSLSLTENYISAQAVOGISNPSG 120

QY 121 DLSSGALGCEPKFENVDETAVTGSGWGRPQRDGPALRATAMI GFGQWLLDNGYSTATDIYW 180  
DB 121 DLSSGALGCEPKFENVDETAVTGSGWGRPQRDGPALRATAMI GFGQWLLDNGYSTATDIYW 180  
QY 181 PLVRNDLSVAQYWNQGYDLMEVNGSSPFTTAVOHRALVBSAFATVGGSCWCDSQ 240  
DB 181 PLVRNDLSVAQYWNQGYDLMEVNGSSPFTTAVOHRALVBSAFATVGGSCWCDSQ 240  
QY 241 APEILCYLQSFMTGSEFLANFDSRSRGKANTLLGSIHTFDEPAACDSTFQCSPRALA 300  
DB 241 APEILCYLQSFMTGSEFLANFDSRSRGKANTLLGSIHTFDEPAACDSTFQCSPRALA 300  
QY 301 NHKEVVDSTFRSYITLNDGLSDSEAVAVGRYPEDTYNGNPMPLCTIAABOLYDALYOND 360  
DB 301 NHKEVVDSTFRSYITLNDGLSDSEAVAVGRYPEDTYNGNPMPLCTIAABOLYDALYOND 360  
QY 361 KQSLLEVTVSDLDFEFALYSDAATGYSSSSSTYSIVDAVKTFADGFSIVETHASNG 420  
DB 361 KQSLLEVTVSDLDFEFALYSDAATGYSSSSSTYSIVDAVKTFADGFSIVETHASNG 420  
QY 421 SMSEQYDKSDGEQLSARDLTWSYALLLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480  
DB 421 SMSEQYDKSDGEQLSARDLTWSYALLLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480  
QY 481 TVSSVTVTSPSIVATGTTTTATPTGSGSVTSTKTTATASKSTTTTSGMSL 534  
DB 481 TVSSVTVTSPSIVATGTTTTATPTGSGSVTSTKTTATASKSTTTTSGMSL 534

## RESULT 2

AA18090 ID AA18090 standard; protein; 534 AA.

XX AA18090;

XX 02-SEP-1999 (first entry)

XX Aspergillus niger glucoamylase enzyme.

XX Glucoamylase enzyme; saccharification; starch hydrolyase;

XX dextrose syrup; syrup; ethanol; fuel; beverage; citric acid;

XX ascorbic acid; lysine; glutamic acid.

XX Aspergillus niger.

XX MO9928448-A1.

XX 10-JUN-1999.

XX 26-NOV-1998; 98MO-DK000520.

XX 26-NOV-1997; 97US-00979673.

XX 30-DEC-1997; 97DK-00001557.

XX 30-JUN-1998; 98US-00107657.

XX 10-JUL-1998; 98DK-00000925.

XX (NOVO ) NOVO-NORDISK AS.

XX Nielsen BR, Nielsen RI, Lehmbeck J;

XX WPI; 1999-404822/34.

XX N-PSDB; AAX81797.

XX New glucoamylase obtained from Talaromyces emersonii.

XX Disclosure; Page 63-64; 79pp; English.

XX The present sequence represents a glucoamylase enzyme obtained from  
XX Aspergillus niger. The specification describes a Talaromyces emersonii  
XX glucoamylase enzyme. The glucoamylase enzymes have high thermal stability  
XX so that a saccharification process may be carried out within a shorter  
XX period of time or the process may be carried out using a lower enzyme  
XX dosage. The glucoamylase enzymes can be used for saccharifying starch

CC hydrolyzate for converting starch or partially hydrolyzed starch into a  
CC syrup containing dextrose. They can be used for producing  
CC oligosaccharides, specially syrups, ethanol for fuel, beverages or  
CC organic compounds such as citric acid, ascorbic acid, lysine or glutamic  
CC acid

SQ Sequence 534 AA;

Query Match 100.0%; Score 2771; DB 2; Length 534;  
Best Local Similarity 100.0%; Pred. No. 1.5e-214;  
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 MSFRLSLALSGVCTGLANVISKRAITLDSMLNEATVARTALINNIGADGAWVSGADSGI 60
DB 1 MSFRLSLALSGVCTGLANVISKRAITLDSMLNEATVARTALINNIGADGAWVSGADSGI 60
OY 61 VVASSTNDPDYFYWTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYQGISNPSG 120
DB 61 VVASSTNDPDYFYWTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYQGISNPSG 120
OY 61 VVASSTNDPDYFYWTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYQGISNPSG 120
DB 61 VVASSTNDPDYFYWTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYQGISNPSG 120
OY 121 DLSGAGLGEPEFNVDETAYTSGWGRPQRDGPALRATAMIGQWMLDNGYSTATDIYW 180
DB 121 DLSGAGLGEPEFNVDETAYTSGWGRPQRDGPALRATAMIGQWMLDNGYSTATDIYW 180
OY 181 PLVRNDLSVYAQYMNQGYDLMEVNGSSFTIAYQHRALVEGSAFATAVSGSCSWCDSQ 240
DB 181 PLVRNDLSVYAQYMNQGYDLMEVNGSSFTIAYQHRALVEGSAFATAVSGSCSWCDSQ 240
OY 241 APEILICYLOSFWTGSFTILANFDSSRSRGKANTLLGSIHFDPEAACDDSTFQPCSPRALA 300
DB 241 APEILICYLOSFWTGSFTILANFDSSRSRGKANTLLGSIHFDPEAACDDSTFQPCSPRALA 300
OY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGMPFLCTLAAEQLYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGMPFLCTLAAEQLYDALYQMD 360
OY 361 KQGSLEVTDSLDFKALYSDAATGTYSSSSSTYSISVDAVKTFFADGFSIVETHAASNG 420
DB 361 KQGSLEVTDSLDFKALYSDAATGTYSSSSSTYSISVDAVKTFFADGFSIVETHAASNG 420
OY 421 SMSBOYKSDGQOLSRDLTWSYALALTANNRRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSBOYKSDGQOLSRDLTWSYALALTANNRRNSVVPASWGETSASVPGTCAATSAIG 480
OY 481 TYSSVTWSPISIVATGTTTATPTGSGSVSTSKTTATASKTSTTTSSGMSL 534
DB 481 TYSSVTWSPISIVATGTTTATPTGSGSVSTSKTTATASKTSTTTSSGMSL 534

RESULT 3
AAB03450 standard; protein; 534 AA.
ID AAB03450
XX
AC AAB03450;
XX
DT 03-JAN-2001 (first entry)
XX
DE Aspergillus niger G2 glucoamylase.
XX
KM Glucoamylase G2; starch hydrolysis; high fructose corn syrup;
XX thermal stability.
XX
OS Aspergillus niger.
XX
FH Key
FH Peptide 1..27
FT /label= prepropeptide_sequence
FT Cleavage-site 27..28
FT /label= tripeptidyl_aminopeptidase_cleavage_site
FT Protein 28..534
FT /label= mature_glucoamylase
XX
PN WO200034452-A1.
```

```
XX 15-JUN-2000.
PD 07-DEC-1999; 99WO-DK000686.
XX
PF 07-DEC-1999; 98DK-00001616.
PR 24-MAR-1999; 99DK-00000409.
XX
PA (NOVO ) NOVO-NORDISK AS.
XX
PI Nielsen BR, Svendsen A, Boisen K, Vind J, Pedersen H;
XX WPI; 2000-431296/37.
DR
XX Variance of parent fungal glucoamylase useful for producing ethanol,
PT beverages and in fermentation processes comprise a peptide extension at
PT the amino terminal.
XX
PS Disclosure; Page 50-51; 61pp; English.
XX
CC The present sequence is the G2 glucoamylase from Aspergillus niger. It
CC was used to construct a version of the enzyme which has improved thermal
CC stability, and which can be used more efficiently in starch hydrolysis.
CC This process is used during high fructose corn syrup production, as well
CC as in the production of ethanol for fuel or beverage, and in fermentation
CC processes for producing organic compounds such as citric acid, ascorbic
CC acid, lysine and glutamic acid
XX
SQ Sequence 534 AA;

Query Match 100.0%; Score 2771; DB 3; Length 534;
Best Local Similarity 100.0%; Pred. No. 1.5e-214;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSFRLSLALSGVCTGLANVISKRAITLDSMLNEATVARTALINNIGADGAWVSGADSGI 60
DB 1 MSFRLSLALSGVCTGLANVISKRAITLDSMLNEATVARTALINNIGADGAWVSGADSGI 60
OY 61 VVASSTNDPDYFYWTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYQGISNPSG 120
DB 61 VVASSTNDPDYFYWTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYQGISNPSG 120
OY 121 DLSGAGLGEPEFNVDETAYTSGWGRPQRDGPALRATAMIGQWMLDNGYSTATDIYW 180
DB 121 DLSGAGLGEPEFNVDETAYTSGWGRPQRDGPALRATAMIGQWMLDNGYSTATDIYW 180
OY 181 PLVRNDLSVYAQYMNQGYDLMEVNGSSFTIAYQHRALVEGSAFATAVSGSCSWCDSQ 240
DB 181 PLVRNDLSVYAQYMNQGYDLMEVNGSSFTIAYQHRALVEGSAFATAVSGSCSWCDSQ 240
OY 241 APEILICYLOSFWTGSFTILANFDSSRSRGKANTLLGSIHFDPEAACDDSTFQPCSPRALA 300
DB 241 APEILICYLOSFWTGSFTILANFDSSRSRGKANTLLGSIHFDPEAACDDSTFQPCSPRALA 300
OY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGMPFLCTLAAEQLYDALYQMD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGMPFLCTLAAEQLYDALYQMD 360
OY 361 KQGSLEVTDSLDFKALYSDAATGTYSSSSSTYSISVDAVKTFFADGFSIVETHAASNG 420
DB 361 KQGSLEVTDSLDFKALYSDAATGTYSSSSSTYSISVDAVKTFFADGFSIVETHAASNG 420
OY 421 SMSBOYKSDGQOLSRDLTWSYALALTANNRRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSBOYKSDGQOLSRDLTWSYALALTANNRRNSVVPASWGETSASVPGTCAATSAIG 480
OY 481 TYSSVTWSPISIVATGTTTATPTGSGSVSTSKTTATASKTSTTTSSGMSL 534
DB 481 TYSSVTWSPISIVATGTTTATPTGSGSVSTSKTTATASKTSTTTSSGMSL 534

RESULT 4
AAY77740
```

ID AAY77740 standard; protein; 534 AA.  
 XX  
 AC AAY77740;  
 XX  
 DT 22-MAY-2000 (first entry)  
 XX  
 DE A. niger G2 glucamylase.  
 XX  
 KW Glucamylase; variant; starch conversion; saccharification; ethanol;  
 KM fuel; beverage; fermentation; citric acid; ascorbic acid; thermostable;  
 XX G2 glucamylase; fungl.  
 XX  
 OS Aspergillus niger.  
 XX  
 PN WO200004136-A1.  
 XX  
 PD 27-JAN-2000.  
 XX  
 PF 09-JUL-1999; 99MO-DK000392.  
 XX  
 PR 15-JUL-1998; 98DK-00000937.  
 PR 17-DEC-1998; 98DK-00001667.  
 XX  
 PA (NOVO ) NOVO-NORDISK AS.  
 XX  
 PI Nielsen BR, Svendsen A, Pedersen H, Vind J, Hendriksen HV;  
 PI Frandsen TP;  
 DR WPI; 2000-182412/16.  
 DR N-PSDB; AAB287842.  
 XX  
 PT Variant fungal glucamylases with improved thermostability and increased  
 PT specific activity, useful in saccharification processes.  
 XX  
 PS Claim 1; Page 78-81; 116pp; English.  
 XX  
 CC The invention relates to variant fungal glucamylases comprising specific  
 CC mutations. The variant comprises one or more mutation in position/region  
 CC 1-18, 19-35, 40-62, 73-80, 93-127, 170-184, 200-212, 234-246, 287-319,  
 CC 334-341, 353-374, 388-414, 445-470 of the parent G2 glucamylase sequence  
 CC (AAY77740) from A. niger, with the exception of: N20C, A27C, S30P, Y48W,  
 CC Y50F, W52F, R54K/L, D55G/V, G57A, K108R, D112Y, Y16A/W, S19C/W/E/G/Y/P,  
 CC W120H/L/F/Y, G121T/A, R122Y, P123G, Q124H, R125K, W170F, N171S, Q172N, CC  
 T173G, G174C, I175F, D176N/E, L177H/D, W178R/D, E179Q/D, E180D/Q,  
 CC V181D/A/T, N182A/D/Q/Y/S, G183K, S184H, W212F, R241K, A246C, D293E/Q,  
 CC A302V, R305K, Y306F, D309N/E, Y312W, W317F, E389D/Q, H391W, A392D, A393P,  
 CC N395Q, G396S, E400C/C, Q401E, G407D, E408P, L410F, S411A/G/C/H/D, and  
 CC S460P. The glucamylase variants are useful in a starch conversion  
 CC process, especially continuous process which include a continuous  
 CC saccharification process. The variants can be used for producing  
 CC oligosaccharides, specialty syrups, or ethanol for fuel or beverages.  
 CC They can also be used in fermentation processes for producing organic  
 CC compounds such as citric acid, ascorbic acid, lysine and glutamic acid.  
 CC The glucamylase variants have improved thermostability and/or increased  
 CC specific activity. This is advantageous in industrial saccharification  
 CC processes. The risk of microbial contamination is also reduced when  
 CC carrying the saccharification process at temperatures above 63 plusOC. An  
 CC increased specific activity towards short chain saccharides such as  
 CC maltose (without reducing the activity towards oligosaccharides) would  
 CC also permit using a lower enzyme dosage and/or shorter process times. The  
 CC present sequence represents the G2 glucamylase from A. niger, the parent  
 CC glucamylase used for constructing the variants  
 XX  
 SQ Sequence 534 AA:

Query Match 100.0%; Score 2771; DB 3; Length 534;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-214;  
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFRSLALSLGLVCTGLIANYISKRATDLSWLSNEATVAFATLANNIGADGAWWSGADSGI 60  
 DB 1 MSFRSLALSLGLVCTGLIANYISKRATDLSWLSNEATVAFATLANNIGADGAWWSGADSGI 60

QY 61 VVASPESTNDPVPFYTTWTRDSGLVLTUVDLFRNGDPTSLSTIENTYISAQAIYOGISNPBG 120  
 DB 61 VVASPESTNDPVPFYTTWTRDSGLVLTUVDLFRNGDPTSLSTIENTYISAQAIYOGISNPBG 120  
 QY 121 DLSSGAGLGEPKFKNVDETAVTGSGWGRPORDDPALRATAMIGFQOMLLDNGYSTATDIYW 180  
 DB 121 DLSSGAGLGEPKFKNVDETAVTGSGWGRPORDDPALRATAMIGFQOMLLDNGYSTATDIYW 180  
 QY 181 PLVRNDLSVYAOYWNQGYDLMEEVNGSSFFTIAVOHRALVEGSAFATVAGSSCWDSQ 240  
 DB 181 PLVRNDLSVYAOYWNQGYDLMEEVNGSSFFTIAVOHRALVEGSAFATVAGSSCWDSQ 240  
 QY 241 APEILCYLOSFWTGSFTILANPDSRSRGKANTLLGSIHFDEPAACDDSTPOCSPRALA 300  
 DB 241 APEILCYLOSFWTGSFTILANPDSRSRGKANTLLGSIHFDEPAACDDSTPOCSPRALA 300  
 QY 301 NHKEVVDSPRSIYTLNDGLDSEAVAVGVRPEDTYVNGNPMFLCTLAALAEQLYDALYQND 360  
 DB 301 NHKEVVDSPRSIYTLNDGLDSEAVAVGVRPEDTYVNGNPMFLCTLAALAEQLYDALYQND 360  
 QY 361 KQGSLEVTDVSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTADGFPVSIYETHAASNG 420  
 DB 361 KQGSLEVTDVSLDFFKALYSDAATGYSSSSSTYSSIVDAVKTADGFPVSIYETHAASNG 420  
 QY 421 SMSEQYDKSDGRQLSARDLTWSYAALLTANNRNSVVPASWGETSASSVPGTCAATSAIG 480  
 DB 421 SMSEQYDKSDGRQLSARDLTWSYAALLTANNRNSVVPASWGETSASSVPGTCAATSAIG 480  
 QY 481 TYSSVTVTSMPSIVATGCTTTTATPTGSGSVTSTKTTATASXTSTYTSAGMSL 534  
 DB 481 TYSSVTVTSMPSIVATGCTTTTATPTGSGSVTSTKTTATASXTSTYTSAGMSL 534

## RESULT 5

ID AAB48171  
 AAB48171 standard; protein; 534 AA.

AC AAB48171;

DT 02-APR-2001 (first entry)

XX A. niger G1 glucamylase polypeptide.

KM Thermosacus crustaceus; glucamylase; starch conversion; ethanol;

KM maltose syrup; beverage; citric acid; ascorbic acid; detergent;

XX thermostability; glucose; G1 glucamylase.

OS Aspergillus niger.

FH Key Location/Qualifiers

FT Peptide 1..24

FT Protein /note="signal peptide"

FT /note="534

FT /note="mature protein"

PN WO200075296-A1.

PD 14-DEC-2000.

PF 02-JUN-2000; 2000MO-DK000301.

PR 02-JUN-1999; 99DK-00000779.

XX (NOVO ) NOVO NORDISK AS.

PI Nielsen BR, Kauppinen MS, Nielsen RI;

PT WPI; 2001-071066/08.

DR N-PSDB; AAC84444.

XX Novel Thermosacus crustaceus glucamylase polypeptide useful in a

PT continuous starch conversion process, detergents, ethanol for fuel or drinking, and

PT oligosaccharides, specialty syrups, ethanol for fuel or drinking, and

PT beverages.

XX Disclosure; Page 52-55; 61pp; English.

CC The invention provides a *Thermascus crustaceus* glucoamylase polypeptide.  
 CC The glucoamylase polypeptide can be expressed by standard recombinant  
 CC methodology and is useful in a continuous starch conversion process, for  
 CC producing oligosaccharides, specialty syrups such as maltose syrup,  
 CC ethanol for fuel or drinking ethanol, beverages, and organic compounds  
 CC such as citric acid, ascorbic acid, lysine or glutamic acid. It is also  
 CC useful in detergent such as laundry detergent compositions, dish wash  
 CC compositions and/or hard surface cleaning compositions. The *T. crustaceus*  
 CC glucoamylase has higher thermostability than *Aspergillus niger* G1  
 CC glucoamylase. It also has higher specific activity and/or decreased  
 CC glucoamylase reversion tendency. The present sequence represents the *A. niger*  
 CC G1 glucoamylase

SQ Sequence 534 AA;

Query Match 100.0%; Score 2771; DB 4; Length 534;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-214;  
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFRLSLALSGVCTGLANVISKRAITLDSMLSEATVARTAILNNIGADGAWSGADSGI 60  
 DB 1 MSFRLSLALSGVCTGLANVISKRAITLDSMLSEATVARTAILNNIGADGAWSGADSGI 60  
 QY 61 VVASPSTNDPDIYFTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIVOGINSPEG 120  
 DB 61 VVASPSTNDPDIYFTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIVOGINSPEG 120  
 QY 121 DISSGAGLGEPPKPNVDETYTSGWGRPORDBPALRTAMIGFQWLLDNGYSTADIYW 180  
 DB 121 DISSGAGLGEPPKPNVDETYTSGWGRPORDBPALRTAMIGFQWLLDNGYSTADIYW 180  
 QY 181 PLVRNDLSVVAQYMNQGYDLMEEVNGSSFFITIAVQHRALVEGSAFATAVSGSCWCDSQ 240  
 DB 181 PLVRNDLSVVAQYMNQGYDLMEEVNGSSFFITIAVQHRALVEGSAFATAVSGSCWCDSQ 240  
 QY 241 APEILCYLOSFWTGSFTILANFDSRSRSGKANTLLGSIHFPDPAACDSTFQPCSPRALA 300  
 DB 241 APEILCYLOSFWTGSFTILANFDSRSRSGKANTLLGSIHFPDPAACDSTFQPCSPRALA 300  
 QY 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPETTYNCPWFLCTLAABQLYDALYQMD 360  
 DB 301 NHKEVVDSPRSIYTLNDGSDSEAVAVGRYPETTYNCPWFLCTLAABQLYDALYQMD 360  
 QY 361 KGSLEVTDLVSLDFKALYSDAATGTYSSSSSTYSIVDAVKTFADGFVSIYETHAASNG 420  
 DB 361 KGSLEVTDLVSLDFKALYSDAATGTYSSSSSTYSIVDAVKTFADGFVSIYETHAASNG 420  
 QY 421 SMSBOYDKSDGQSLARDLTWYSYALLLTANNRRNSVVPASWGETSASSVPGTCAATSAIG 480  
 DB 421 SMSBOYDKSDGQSLARDLTWYSYALLLTANNRRNSVVPASWGETSASSVPGTCAATSAIG 480  
 QY 481 TYSSVTVTSPISVATGCTTTTATPTGSGSVSTSTTTATASTSTTTTSGMWL 534  
 DB 481 TYSSVTVTSPISVATGCTTTTATPTGSGSVSTSTTTATASTSTTTTSGMWL 534

RESULT 6

AAB61904  
 ID AAB61904 standard; protein; 534 AA.

AC AAB61904;

DT 08-MAY-2001 (first entry)

DE *A. niger* G2 glucoamylase.

KX Glucoamylase; variant; starch; dextrose; saccharification; maltodextrin;  
 KW glucose syrup; fuel; ethanol; beverage; fermentation; citric acid;  
 KM ascorbic acid.

XX *Aspergillus niger*.

OS Key-<sup>1</sup> Location/Qualifiers  
 FH Peptide 1..24  
 FT Protein /note="signal peptide"  
 FT Protein 25..534  
 FT Protein /note="mature protein"

XX MO200104273-A2.

XX 18-JAN-2001.

XX 07-JUL-2000; 2000MO-DK000373.

XX 09-JUL-1999; 99DK-00000999.

XX (NOVO) NOVO NORDISK AS.

XX Nielsen BR, Svendsen A, Pedersen H, Vind J, Hendriksen HV;

XX Frandsen TP;

XX WPI; 2001-138334/14.

XX N-PSDB; AAC85099.

PT Novel variant of parent glucoamylase useful in starch conversion process,  
 PT and for producing oligosaccharides, maltodextrins, glucose syrups, fuel,  
 PT drinking ethanol, beverage and organic compounds.

PS Claim 1; Page 51-53; 58pp; English.

CC The invention relates to a variant of a parent glucoamylase, comprising  
 CC an alteration at positions 59, 66, 72, 119, 189, 223, 227, 313, 340, 342,  
 CC 352, 379, 386, 393, 395, 402, 408, 416, 425, 427, 444, 486, 490, 494,  
 CC where the alteration is independently an insertion, substitution or  
 CC deletion of amino acid which occupies the position. The variant  
 CC glucoamylase is useful for converting starch or partially hydrolyzed  
 CC starch into a syrup containing dextrose, by saccharifying starch  
 CC hydrolyzate. The variant is useful in the starch conversion process, for  
 CC producing oligosaccharides, maltodextrins or glucose syrups, fuel,  
 CC drinking ethanol, beverage and in a fermentation process for producing  
 CC organic compounds, such as citric acid, ascorbic acid, lysine and  
 CC glutamic acid. It is useful for improving the thermal stability and/or  
 CC specific activity of a parent glucoamylase. The present sequence  
 CC represents an *A. niger* G2 glucoamylase, the parent enzyme from which the  
 CC variant of the invention is derived

SQ Sequence 534 AA;

Query Match 100.0%; Score 2771; DB 4; Length 534;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-214;  
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSFRLSLALSGVCTGLANVISKRAITLDSMLSEATVARTAILNNIGADGAWSGADSGI 60

DB 1 MSFRLSLALSGVCTGLANVISKRAITLDSMLSEATVARTAILNNIGADGAWSGADSGI 60

QY 61 VVASPSTNDPDIYFTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIVOGINSPEG 120

DB 61 VVASPSTNDPDIYFTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIVOGINSPEG 120

QY 121 DISSGAGLGEPPKPNVDETYTSGWGRPORDBPALRTAMIGFQWLLDNGYSTADIYW 180

DB 121 DISSGAGLGEPPKPNVDETYTSGWGRPORDBPALRTAMIGFQWLLDNGYSTADIYW 180

QY 181 PLVRNDLSVVAQYMNQGYDLMEEVNGSSFFITIAVQHRALVEGSAFATAVSGSCWCDSQ 240

DB 181 PLVRNDLSVVAQYMNQGYDLMEEVNGSSFFITIAVQHRALVEGSAFATAVSGSCWCDSQ 240

QY 241 APEILCYLOSFWTGSFTILANFDSRSRSGKANTLLGSIHFPDPAACDSTFQPCSPRALA 300

DB 241 APEILCYLOSFWTGSFTILANFDSRSRSGKANTLLGSIHFPDPAACDSTFQPCSPRALA 300

QY 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABOLYDALYQWD 360  
 Db 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABOLYDALYQWD 360  
 QY 361 KQGSLEVTQVSLDFPFALYSDAATGTSSSSSTYSIVDAVKTFADGFSIVETHAASNG 420  
 Db 361 KQGSLEVTQVSLDFPFALYSDAATGTSSSSSTYSIVDAVKTFADGFSIVETHAASNG 420  
 QY 421 SMEQYDKSDGEQLSARDLTWSYAALLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480  
 Db 421 SMEQYDKSDGEQLSARDLTWSYAALLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480  
 QY 481 TYSSVTVTSMPSTIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTRSGMSL 534  
 Db 481 TYSSVTVTSMPSTIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTRSGMSL 534  
 RESULT 7  
 AAY77741  
 ID AAY77741 standard; protein; 640 AA.  
 XX  
 AC AAY77741;  
 XX  
 DT 22-MAY-2000 (first entry)  
 XX  
 DE A. niger G1 glucoamylase.  
 XX  
 KW Glucoamylase; variant; starch conversion; saccharification; ethanol;  
 KM fuel; beverage; fermentation; citric acid; ascorbic acid; thermostable;  
 KM G1 glucoamylase; fungal.  
 XX  
 OS Aspergillus niger.  
 PN W0200004136-A1.  
 PD 27-JAN-2000.  
 XX  
 PF 09-JUL-1999; 99NO-DK000392.  
 XX  
 PR 15-JUL-1998; 98DK-00000937.  
 PR 17-DEC-1998; 98DK-00001667.  
 XX  
 PA (NOVO ) NOVO-NORDISK AS.  
 XX  
 PI Nielsen BR, Svendsen A, Pedersen H, Vind J, Hendriksen HV;  
 PI Frandsen TP;  
 XX  
 DR N-PSDB; AAZ87843.  
 DR WPI; 2000-182412/16.  
 XX  
 PT Variant fungal glucoamylases with improved thermostability and increased  
 PT specific activity, useful in saccharification processes.  
 XX  
 PS Disclosure; Page 91-93; 116pp; English.  
 XX  
 CC The invention relates to variant fungal glucoamylases. The variants  
 CC comprise specific mutations in the patent G2 glucoamylase (AmG) sequence  
 CC (AAY77740) from A. niger (see AAZ87842 for specific positions of the  
 CC mutations). The glucoamylase variants are useful in a starch conversion  
 CC process, especially continuous process which include a continuous  
 CC saccharification process. The variants can be used for producing  
 CC oligosaccharides, specially syrups, or ethanol for fuel or beverages.  
 CC They can also be used in fermentation processes for producing organic  
 CC compounds such as citric acid, ascorbic acid, lysine and glutamic acid.  
 CC The glucoamylase variants have improved thermostability and/or increased  
 CC specific activity. This is advantageous in industrial saccharification  
 CC processes. The risk of microbial contamination is also reduced when  
 CC carrying the saccharification process at temperatures above 63 plusC. An  
 CC increased specific activity towards short chain saccharides such as  
 CC maltose (without reducing the activity towards oligosaccharides) would  
 CC also permit using a lower enzyme dosage and/or shorter process times. The  
 CC present sequence represents the G1 glucoamylase from A. niger

SEQ Sequence 640 AA:  
 Query Match 99.0%; Score 2742; DB 3; Length 640;  
 Best Local Similarity 99.6%; Pred. No. 4,4e-212;  
 Matches 528; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MSFRSLATSGVCTGLANVISKRATLDSMLSEATVARTAILNNIGADGAWSGADSGI 60  
 Db 1 MSFRSLATSGVCTGLANVISKRATLDSMLSEATVARTAILNNIGADGAWSGADSGI 60  
 QY 61 VVASPETDNPDEYFTYTRDSGLVLTVDLFRNGDTSLSTLENYISAQAIYOGISNPSG 120  
 Db 61 VVASPETDNPDEYFTYTRDSGLVLTVDLFRNGDTSLSTLENYISAQAIYOGISNPSG 120  
 QY 121 DLSSGAGLGPKNVDETAYTSGWGPORDGPALRTAMIGFGQWLLDNGYSTARDIYW 180  
 Db 121 DLSSGAGLGPKNVDETAYTSGWGPORDGPALRTAMIGFGQWLLDNGYSTARDIYW 180  
 QY 181 PLVRNDLSYVAQYMNQGYDIMEVNGSFFITAVOHRALVSGSARFATVSGSCWCDGQ 240  
 Db 181 PLVRNDLSYVAQYMNQGYDIMEVNGSFFITAVOHRALVSGSARFATVSGSCWCDGQ 240  
 QY 241 APEILCYLQSFMTGSFTLANFDSRSRSGKDANTLLGSIHTFDPACDSDTFQCSFRALA 300  
 Db 241 APEILCYLQSFMTGSFTLANFDSRSRSGKDANTLLGSIHTFDPACDSDTFQCSFRALA 300  
 QY 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABOLYDALYQWD 360  
 Db 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABOLYDALYQWD 360  
 QY 361 KQGSLEVTQVSLDFPFALYSDAATGTSSSSSTYSIVDAVKTFADGFSIVETHAASNG 420  
 Db 361 KQGSLEVTQVSLDFPFALYSDAATGTSSSSSTYSIVDAVKTFADGFSIVETHAASNG 420  
 QY 421 SMEQYDKSDGEQLSARDLTWSYAALLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480  
 Db 421 SMEQYDKSDGEQLSARDLTWSYAALLTANNRRNSVVPASWGETSASVPGTCAATSAIG 480  
 QY 481 TYSSVTVTSMPSTIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTRSGMSL 530  
 Db 481 TYSSVTVTSMPSTIVATGCTTTTATPTGSGSVTSTSKTTATASKTSTTRSGMSL 530  
 RESULT 8  
 AAB61905  
 ID AAB61905 standard; protein; 640 AA.  
 XX  
 AC AAB61905;  
 XX  
 DT 08-MAY-2001 (first entry)  
 XX  
 DE A. niger protein sequence id No. 13.  
 XX  
 KW Glucoamylase; variant; starch; dextrose; saccharification; maltodextrin;  
 KM glucose syrup; fuel; ethanol; beverage; fermentation; citric acid;  
 KM ascorbic acid.  
 XX  
 OS Aspergillus niger.  
 PN W0200104273-A2.  
 PD 18-JAN-2001.  
 XX  
 PF 07-JUL-2000; 2000MO-DK000373.  
 XX  
 PR 09-JUL-1999; 99DK-00000999.  
 XX  
 PA (NOVO ) NOVO-NORDISK AS.  
 XX  
 PI Nielsen BR, Svendsen A, Pedersen H, Vind J, Hendriksen HV;  
 PI Frandsen TP;  
 XX  
 DR WPI; 2001-138334/14.



XX Novel variant of parent glucoamylase useful in starch conversion process,  
 PT and for producing oligosaccharides, maltodextrins, glucose syrups, fuel,  
 PT drinking ethanol, beverage and organic compounds.

XX Disclosure, Page 57-58; 58pp; English.

XX The invention relates to a variant of a parent glucoamylase, comprising  
 CC an alteration at positions 59, 66, 72, 119, 189, 223, 227, 313, 340, 342,  
 CC 352, 379, 386, 393, 395, 402, 408, 416, 425, 444, 486, 490, 494,  
 CC where the alteration is independently an insertion, substitution or  
 CC deletion of amino acid which occupies the position. The variant  
 CC glucoamylase is useful for converting starch or partially hydrolyzed  
 CC starch into a syrup containing dextrose, by saccharifying starch  
 CC hydrolyzate. The variant is useful in the starch conversion process, for  
 CC producing oligosaccharides, maltodextrins or glucose syrups, fuel,  
 CC drinking ethanol, beverage and in a fermentation process for producing  
 CC organic compounds, such as citric acid, ascorbic acid, lysine and  
 CC glutamic acid. It is useful for improving the thermal stability and/or  
 CC specific activity of a parent glucoamylase

XX Sequence 640 AA;

Query Match 99.0%; Score 2742; DB 4; Length 640;  
 Best Local Similarity 99.6%; Pred. No. 4.4e-212;  
 Matches 528; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 MSFRLSLALSGVCTGLANVISKRATLDWSLNEAVARTAILNINIGAGAWVSGADSGI 60
DB 1 MSFRLSLALSGVCTGLANVISKRATLDWSLNEAVARTAILNINIGAGAWVSGADSGI 60
QY 61 VVASPTDNDPDIFFYTWTRDSGLVLTVDLFRNGDTSLSSTIENTYISAQAIYVGISNPSG 120
DB 61 VVASPTDNDPDIFFYTWTRDSGLVLTVDLFRNGDTSLSSTIENTYISAQAIYVGISNPSG 120
QY 121 DISSGAGLGEPRFNVDETAVTGSMGRPQRDGPALRATAMIGFQWLLDNGYSTATDIYW 180
DB 121 DISSGAGLGEPRFNVDETAVTGSMGRPQRDGPALRATAMIGFQWLLDNGYSTATDIYW 180
QY 181 PLVRNDLSVAQYVMOTGYDLMEEVNNGSFFITAVGHRALVGSAAATVAGSSCSWCDSQ 240
DB 181 PLVRNDLSVAQYVMOTGYDLMEEVNNGSFFITAVGHRALVGSAAATVAGSSCSWCDSQ 240
QY 241 APEILCYLOSFWTGSFIILANFDSRSRGKANTLLGSIHFDPPEACDDSTFQPCSRALA 300
DB 241 APEILCYLOSFWTGSFIILANFDSRSRGKANTLLGSIHFDPPEACDDSTFQPCSRALA 300
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABEQLYDALYQWD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABEQLYDALYQWD 360
QY 361 KQGSLEVDVSLDFFKALYSDAATGTYSSTSSSTYSIVDAVKTFAAGFYSIVETHAASNG 420
DB 361 KQGSLEVDVSLDFFKALYSDAATGTYSSTSSSTYSIVDAVKTFAAGFYSIVETHAASNG 420
QY 421 SMSBOYDKSDGQOLSRDLTWGVAALTLNANRNSVVPASWGETSASVPGTCAATSAIG 480
DB 421 SMSBOYDKSDGQOLSRDLTWGVAALTLNANRNSVVPASWGETSASVPGTCAATSAIG 480
QY 481 TVSSVTVTWSPSIVATGTTTATPTGSGSVTSTKTTATASKTSTTTS 530
DB 481 TVSSVTVTWSPSIVATGTTTATPTGSGSVTSTKTTATASKTSTTTS 530

```

RESULT 9  
 AAP40212  
 ID AAP40212 standard; protein; 639 AA.

AC AAP40212;  
 XX 25-MAR-2003 (revised)  
 DT 09-JAN-1992 (first entry)  
 XX

DE Sequence encoded by A.awamori glucoamylase genomic region.  
 XX Starch.hydrolysis; glucose.

XX Aspergillus awamori.

XX WO8402921-A.

XX 02-AUG-1984.

XX 26-JUN-1984; 84WO-US000122.

XX 28-JAN-1983; 83US-00461920.

XX 20-DEC-1983; 83US-00563941.

XX 27-APR-1987; 87US-00047552.

XX (CETU ) CETUS CORP.

XX (CETU ) CETUS CORP.

XX Nunberg JH, Flatgaard JE, Innis MA, Gelfand DH, Meade JH;

XX WPI, 1984-201413/32.

XX N-PSDB; AAN40166.

XX DNA sequence coding for fungal glucoamylase protein - for expression in  
 PT yeast etc. for prodn. of the enzyme.

XX Example; Table I, Page 21-25; 66pp; English.

XX The inventors claim a modified DNA sequence coding for fungal  
 CC glucoamylase protein or its single or multiple base substitutions  
 CC deletions, insertions or inversions is new (see AAN40165). It is derived  
 CC from natural, synthetic or semisynthetic sources and is capable, when  
 CC correctly combined with a cleaved expression vector, of expressing a non-  
 CC native protein having glucoamylase activity on transformation of a host  
 CC organism by the vector. (Updated on 25-MAR-2003 to correct PI field.)

XX Sequence 639 AA;

Query Match 96.8%; Score 2663.5; DB 1; Length 639;  
 Best Local Similarity 98.1%; Pred. No. 2.3e-207;  
 Matches 520; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

```

QY 1 MSFRLSLALSGVCTGLANVISKRATLDWSLNEAVARTAILNINIGAGAWVSGADSGI 60
DB 1 MSFRLSLALSGVCTGLANVISKRATLDWSLNEAVARTAILNINIGAGAWVSGADSGI 60
QY 61 VVASPTDNDPDIFFYTWTRDSGLVLTVDLFRNGDTSLSSTIENTYISAQAIYVGISNPSG 120
DB 61 VVASPTDNDPDIFFYTWTRDSGLVLTVDLFRNGDTSLSSTIENTYISAQAIYVGISNPSG 120
QY 121 DISSGAGLGEPRFNVDETAVTGSMGRPQRDGPALRATAMIGFQWLLDNGYSTATDIYW 180
DB 121 DISSGAGLGEPRFNVDETAVTGSMGRPQRDGPALRATAMIGFQWLLDNGYSTATDIYW 180
QY 181 PLVRNDLSVAQYVMOTGYDLMEEVNNGSFFITAVGHRALVGSAAATVAGSSCSWCDSQ 240
DB 181 PLVRNDLSVAQYVMOTGYDLMEEVNNGSFFITAVGHRALVGSAAATVAGSSCSWCDSQ 240
QY 241 APEILCYLOSFWTGSFIILANFDSRSRGKANTLLGSIHFDPPEACDDSTFQPCSRALA 300
DB 241 APEILCYLOSFWTGSFIILANFDSRSRGKANTLLGSIHFDPPEACDDSTFQPCSRALA 300
QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABEQLYDALYQWD 360
DB 301 NHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNNGNPMFLCTLAABEQLYDALYQWD 360
QY 361 KQGSLEVDVSLDFFKALYSDAATGTYSSTSSSTYSIVDAVKTFAAGFYSIVETHAASNG 420
DB 361 KQGSLEVDVSLDFFKALYSDAATGTYSSTSSSTYSIVDAVKTFAAGFYSIVETHAASNG 420
QY 421 SMSBOYDKSDGQOLSRDLTWGVAALTLNANRNSVVPASWGETSASVPGTCAATSAIG 480

```

```
Db      420 SMSBQYDKSDGDEQLSARDLTWSYAALLTANRRNSVVPASWGERTSASVPGTCAATSAIG 479
Qy      481 TVSSVTVTSMPSIVATGCTTTTATPTGSGSVTSTSKTTATASXTSTTTS 530
Db      480 TVSSVTVTSMPSIVATGCTTTTATPTGSGSVTSTSKTTATASXTSTTTS 529

RESULT 10
AAP81876
ID      AAP81876 standard; protein; 630 AA.
XX
AC      AAP81876;
XX
DT      24-OCT-2003 (revised)
DT      25-MAR-2003 (revised)
DT      15-MAR-1992 (first entry)
XX
DE      Sequence of amyloglucosidase.
XX
KW      Enzyme; brewing; bread-making; dextrin.
XX
OS      Aspergillus niger; NCI 22343.
XX
PN      EP260160-A.
XX
PD      16-MAR-1988.
XX
PF      10-JUN-1987; 87EP-00401300.
XX
PR      10-JUN-1986; 86PR-00008387.
PR      13-APR-1987; 87PR-00005207.
PR      13-APR-1987; 87PR-00005208.
XX
PA      (TRGE ) TRANSGENE SA.
XX
PI      Labat N, Loison G, Lemoine Y;
XX
DR      WPI; 1988-072593/11.
DR      N-PSDB; AAN82019.
XX
PT      New DNA block for expressing amyloglucosidase in yeast - contg. gene plus
PT      transcription and export sequences, and transformed cells useful in
PT      brewing and bread making.
XX
PS      Example; Fig 2; 38pp; French.
XX
CC      EcoRI-Sal I fragments of Aspergillus niger NCI 22343 were cloned in
CC      pBR32 and two clones contg. portions of the gene isolated using a pool
CC      of three probes-TG282, TG283, TG284- (AAN82014-6) designed on the basis
CC      of already published SQs. The clones were designated pTG1830 and 1831. A
CC      CDNA bank of A.niger NCI 22343 was screened for amyloglucosidase using
CC      probe TG33 (AAN82017) which corresp. to AAs 163-172 of the mature
CC      protein (see AAN82019). Oligo TG387 (AAN82018) was used to confirm the
CC      identity of the clones detected (Updated on 25-MAR-2003 to correct PR
CC      field.) (Updated on 25-MAR-2003 to correct PA field.) (Updated on 24-OCT-
CC      2003 to standardise OS field)
XX
SQ      Sequence 630 AA;

Query Match      95.9%; Score 2658; DB 1; Length 630;
Best Local Similarity 97.2%; Pred. No. 2,6e-205;
Matches 515; Conservative 2; Mismatches 3; Indels 10; Gaps 1;
```

```
Db      121 DUSSGAGLGEPRKRVNDEIATYTGSMGRPORD-----GGQWLLDNGYTTAIDIW 170
Qy      181 PLVRNDLSVVAQYWNQGYDLWEVNGSSFFTIAYQHRALVEGSAFATVAGSSCSWCD9Q 240
Db      171 PLVRNDLSVVAQYWNQGYDLWEVNGSSFFTIAYQHRALVEGSAFATVAGSSCSWCD9Q 230
Qy      241 APETILCYLOSFWTGSITLANFDSNRGKQDANTLGSIHFPDPPAACDDSTFOCSPRALA 300
Db      231 APETILCYLOSFWTGSITLANFDSNRGKQDANTLGSIHFPDPPAACDDSTFOCSPRALA 290
Qy      301 NHEKVVDSFRSIYTLNDGSDSEAVAVGRYPEDTYNGNPMWFLCTLAABEQLDALYQMD 360
Db      291 NHEKVVDSFRSIYTLNDGSDSEAVAVGRYPEDTYNGNPMWFLCTLAABEQLDALYQMD 350
Qy      361 KQGSLEVTDSLDFFKALYSDAATGYSSSSSTYSIVDAVKTFFADGFVSIYETHAASNG 420
Db      351 KQGSLEVTDSLDFFKALYSDAATGYSSSSSTYSIVDAVKTFFADGFVSIYETHAASNG 410
Qy      421 SMSBQYDKSDGDEQLSARDLTWSYAALLTANRRNSVVPASWGERTSASVPGTCAATSAIG 480
Db      411 SMSBQYDKSDGDEQLSARDLTWSYAALLTANRRNSVVPASWGERTSASVPGTCAATSAIG 470
Qy      481 TVSSVTVTSMPSIVATGCTTTTATPTGSGSVTSTSKTTATASXTSTTTS 530
Db      471 TVSSVTVTSMPSIVATGCTTTTATPTGSGSVTSTSKTTATASXTSTTTS 520

RESULT 11
AAB15176
ID      AAB15176 standard; protein; 616 AA.
XX
AC      AAB15176;
XX
DT      11-DEC-2000 (first entry)
XX
DE      Aspergillus awamori glucamylase.
XX
KW      Glucoamylase; enzyme; carbohydrase; glucose;
KW      1,4-alpha-D-glucan glucohydrolase.
XX
OS      Aspergillus awamori.
XX
PN      WO200043504-A1.
XX
PD      27-JUL-2000.
XX
PF      10-JAN-2000; 2000MO-US000532.
XX
PR      22-JAN-1999; 99US-00236063.
XX
PA      (IOWA ) UNIV IOWA STATE RES FOUND INC.
XX
PI      Allen MJ, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzatko R;
PI      Ford C;
XX
DR      WPI; 2000-514725/46.
XX
PT      Fungal glucamylase for selective production of glucose rather than alpha
PT      -1,6 linked disaccharide isomaltose, has mutation pair Asn20Cys coupled
PT      with Ala27Cys forming disulfide bond between the two stabilizing members.
XX
PS      Disclosure; Page 152-153; 160pp; English.
XX
CC      The present sequence is glucamylase (1,4-alpha-D-glucan glucohydrolase;
CC      E.C. 3.2.1.3). This enzyme is a carbohydrase, attacking alpha-(1,4)-,
CC      the nonreducing ends of maltooligosaccharides, attacking alpha-(1,4)-,
CC      and alpha-(1,6)-glucosidic bonds. The present invention relates to mutant
CC      glucamylases (see AAB15178-B15184), which have increased
CC      thermostability, increased pH optimum and reduced isomaltose formation.
CC      The mutant proteins are useful for the selective production of glucose
CC      rather than alpha-1,6 linked disaccharide isomaltose
```

SQ Sequence 616 AA;

Query Match 94.8%; Score 2628; DB 3; Length 616;  
Best Local Similarity 99.6%; Pred. No. 6.5e-203;  
Matches 504; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
QY 25 ATLDLWLSNEATVARTAIINNIIGADGAWYSGADSGIVVAPSTDNPDYTYWTRDSGLV 84
DB 1 ATLDLWLSNEATVARTAIINNIIGADGAWYSGADSGIVVAPSTDNPDYTYWTRDSGLV 60
QY 85 KTLVDFRNGDPSLSLSTENYISAQAIYVGIINPSGDLSSGAGLGPKNVDETAYTGSW 144
DB 61 KTLVDFRNGDPSLSLSTENYISAQAIYVGIINPSGDLSSGAGLGPKNVDETAYTGSW 120
QY 145 GRPORDPALRATAMIGFGOMLNDNGYSTATDIWPLVRNDSYVAQYMNQGYDLME 204
DB 121 GRPORDPALRATAMIGFGOMLNDNGYSTATDIWPLVRNDSYVAQYMNQGYDLME 180
QY 205 VNGSSEFTTAVOHRALVEGSAFATAVGSSCSWCDQAPEILCYLOSFWTGSFILANFDS 264
DB 181 VNGSSEFTTAVOHRALVEGSAFATAVGSSCSWCDQAPEILCYLOSFWTGSFILANFDS 240
QY 265 RSGKDANTLLGSIHTFDPEAACDSTFOPCSPRALANHKEVDSFRSIYTLNDGLSDSA 324
DB 241 RSGKDANTLLGSIHTFDPEAACDSTFOPCSPRALANHKEVDSFRSIYTLNDGLSDSA 300
QY 325 VAVGRYPEDTYNNGNMFPLCTLAABEQLYDALYQMDKQSLKLVTVSLDFKALYSDAAT 384
DB 301 VAVGRYPEDTYNNGNMFPLCTLAABEQLYDALYQMDKQSLKLVTVSLDFKALYSDAAT 360
QY 385 GTYSSSSSTYSSIVDAVKTFADGFVSIETHAASNGSMSEQYDKSDGEOLASADLTWSYA 444
DB 361 GTYSSSSSTYSSIVDAVKTFADGFVSIETHAASNGSMSEQYDKSDGEOLASADLTWSYA 420
QY 445 ALLTANNRNSVVPASWGETSASVPGTCAATSAIGTYSVTTSMPISVATGCTTTAT 504
DB 421 ALLTANNRNSVVPASWGETSASVPGTCAATSAIGTYSVTTSMPISVATGCTTTAT 480
QY 505 PTGSGSVTSTKTTATASKTSTTS 530
DB 481 PTGSGSVTSTKTTATASKTSTTS 506
```

RESULT 12

AAW55979  
ID AAW55979 standard; protein; 616 AA.

AC AAW55979;

DT 27-JUL-1998 (first entry)

DE Aspergillus awamori glucoamylase mutant S411A.

KW Aspergillus awamori; glucoamylase; Aspergillus sp; mutant; fungal; food;

KM fructose; corn; sweetener; 1,4-alpha-D-glucan glucosyltransferase;

XX genetic engineering.

XX Synthetic.

OS Aspergillus awamori.

PN WO9803639-A1.

PF 24-JUL-1997; 97WO-US012983.

PR 24-JUL-1996; 96US-0022578P.

XX 02-AUG-1996; 96US-0023077P.

PI (IOWA ) UNIV IOWA STATE RES FOUND INC.

PI Allen M, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzatko R, Ford C;

XX WPI; 1998-120764/11.

XX Genetically engineered fungal glucoamylase - useful in, e.g. food  
PT industry for production of high fructose corn sweeteners.

XX Claim 10; Page; 97pp; English.

PS The present sequence represents a specifically claimed mutant  
CC glucoamylase from *Aspergillus awamori* (1,4-alpha-D-glucan  
CC glucosyltransferase). The present invention describes fungal glucoamylases  
CC (FG) comprising: a mutation pair Asn20Cys coupled with Ala27Cys forming a  
CC disulphide bond between the 2 members of the pair; and a 311-314loop or  
CC Ser114A mutation. FG can be used in industry for the production of high  
CC fructose corn sweeteners, while the glucose produced by glucoamylase can  
CC be crystallised or used in fermentation to produce organic products, e.g.  
CC citric acid, ascorbic acid, lysine, glutamic acid or ethanol for  
CC beverages and fuel. The mutations provide increased thermal stability,  
CC reduced isomaltose formation and increased pH optimum. N.B. The present  
CC sequence is not given in the specification but is derived from SEQ ID  
XX NO:1 as stated in the claim

SQ Sequence 616 AA;

Query Match 94.7%; Score 2625; DB 2; Length 616;  
Best Local Similarity 99.4%; Pred. No. 1.1e-202;  
Matches 503; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
QY 25 ATLDLWLSNEATVARTAIINNIIGADGAWYSGADSGIVVAPSTDNPDYTYWTRDSGLV 84
DB 1 ATLDLWLSNEATVARTAIINNIIGADGAWYSGADSGIVVAPSTDNPDYTYWTRDSGLV 60
QY 85 KTLVDFRNGDPSLSLSTENYISAQAIYVGIINPSGDLSSGAGLGPKNVDETAYTGSW 144
DB 61 KTLVDFRNGDPSLSLSTENYISAQAIYVGIINPSGDLSSGAGLGPKNVDETAYTGSW 120
QY 145 GRPORDPALRATAMIGFGOMLNDNGYSTATDIWPLVRNDSYVAQYMNQGYDLME 204
DB 121 GRPORDPALRATAMIGFGOMLNDNGYSTATDIWPLVRNDSYVAQYMNQGYDLME 180
QY 205 VNGSSEFTTAVOHRALVEGSAFATAVGSSCSWCDQAPEILCYLOSFWTGSFILANFDS 264
DB 181 VNGSSEFTTAVOHRALVEGSAFATAVGSSCSWCDQAPEILCYLOSFWTGSFILANFDS 240
QY 265 RSGKDANTLLGSIHTFDPEAACDSTFOPCSPRALANHKEVDSFRSIYTLNDGLSDSA 324
DB 241 RSGKDANTLLGSIHTFDPEAACDSTFOPCSPRALANHKEVDSFRSIYTLNDGLSDSA 300
QY 325 VAVGRYPEDTYNNGNMFPLCTLAABEQLYDALYQMDKQSLKLVTVSLDFKALYSDAAT 384
DB 301 VAVGRYPEDTYNNGNMFPLCTLAABEQLYDALYQMDKQSLKLVTVSLDFKALYSDAAT 360
QY 385 GTYSSSSSTYSSIVDAVKTFADGFVSIETHAASNGSMSEQYDKSDGEOLASADLTWSYA 444
DB 361 GTYSSSSSTYSSIVDAVKTFADGFVSIETHAASNGSMSEQYDKSDGEOLASADLTWSYA 420
QY 445 ALLTANNRNSVVPASWGETSASVPGTCAATSAIGTYSVTTSMPISVATGCTTTAT 504
DB 421 ALLTANNRNSVVPASWGETSASVPGTCAATSAIGTYSVTTSMPISVATGCTTTAT 480
QY 505 PTGSGSVTSTKTTATASKTSTTS 530
DB 481 PTGSGSVTSTKTTATASKTSTTS 506
```

RESULT 13

AAAB15180  
ID AAAB15180 standard; protein; 616 AA.

AC AAAB15180;

DT 11-DEC-2000 (first entry)

Accession	Location/Qualifiers
DE	Aspergillus awamori mutant glucosylase S411a substitution.
XX	Glucosylase; enzyme; carbohydrase; glucose;
KW	1,4-alpha-D-glucan glucosylase; mutain; mutation.
XX	Aspergillus awamori.
OS	Synthetic.
XX	Key
XX	Misc-difference 411
XX	/note= "Wild-type Ser substituted by Ala"
XX	WO200043504-A1.
XX	27-JUL-2000.
XX	10-JAN-2000; 2000MO-US000532.
XX	22-JAN-1999; 99US-00236063.
XX	(IOWA ) UNIV IOWA STATE RES FOUND INC.
XX	Allen MJ, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzacko R,
XX	Forde C;
XX	WPI; 2000-514725/46.
XX	Fungal glucosylase for selective production of glucose rather than alpha
XX	-1,6 linked disaccharide isomaltose, has mutation pair Asn20Cys coupled
XX	with Ala27Cys forming disulfide bond between the two stabilizing members.
XX	Claim 10; Page; 160pp; English.
XX	Glucosylase (1,4-alpha-D-glucan glucosylase; E.C. 3.2.1.3) is a
XX	carbohydrase. This enzyme cleaves D-glucose from the nonreducing ends of
XX	maltotriose, maltotetraose, and maltopentaose, attacking alpha-(1,4)-, and alpha-(1,6)-glucosidic
XX	bonds. The present invention relates to mutant glucosylases, which have
XX	increased thermostability, increased pH optimum and reduced isomaltose
XX	formation. The mutant proteins are useful for the selective production of
XX	glucose rather than alpha-1,6 linked disaccharide isomaltose. The present
XX	sequence is one such mutant enzyme. Note: The present sequence is not
XX	shown in the specification but is derived from the Aspergillus awamori
XX	wild-type glucosylase sequence given in pages 152-153 of the sequence
XX	listing (SEQ ID 1)
XX	Sequence 616 AA;
XX	Query Match 94.7%; Score 2625; DB 3; Length 616;
XX	Best Local Similarity 99.4%; Pred. No. 1.1e-202;
XX	Matches 503; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY	25 ATLDSEMLENATVATTAIILNNIGADGAWVSGADSGIIVASPEIDNDYFYTWTRDSGLV 84
DB	1 ATLDSEMLENATVATTAIILNNIGADGAWVSGADSGIIVASPEIDNDYFYTWTRDSGLV 60
QY	85 KTLVDLFNNGDTSLLSTIENYISAQAIYVQGISNPSGDLSSGAGLGEPEKNVDEAYTGSW 144
DB	61 KTLVDLFNNGDTSLLSTIENYISAQAIYVQGISNPSGDLSSGAGLGEPEKNVDEAYTGSW 120
QY	145 GRPQDGDALRATAMIGFGQMLDNGYISTATDIWPLVLRNDLSVAQYMNQTYGLDLEE 204
DB	121 GRPQDGDALRATAMIGFGQMLDNGYISTATDIWPLVLRNDLSVAQYMNQTYGLDLEE 180
QY	205 VNGSSFPTIAVOHRLVBSGAFATVVGSSCSMDQSDAPILCYLQSPFMTGSIITLNPSS 264
DB	181 VNGSSFPTIAVOHRLVBSGAFATVVGSSCSMDQSDAPILCYLQSPFMTGSIITLNPSS 240
QY	265 RSGKANTLIGSIHTFDEPAACDSDTPOCSFRALANKEVVDSESRSYITLNDGLSDEA 324
DB	241 RSGKANTLIGSIHTFDEPAACDSDTPOCSFRALANKEVVDSESRSYITLNDGLSDEA 300
QY	325 VAVGRYEDDTYNGNPMFLCTLAEEQLYDALYQMDKQSGLEVTVDLSDFEFALYSDAAT 384

Query Match	94..4%	Score 2615;	DB 2;	Length 616;
Db	301	VAVGRPEDDTYNGNWFPLCTLLAAAEQLYDLALYQMDKGSLEVTVDLDPFKALYSDAAT	360	
Qy	385	GTYSSTSSSTYSIVDAVKTFADGPFVSIYETHAASNGSMSEQYDKSDGEQLSARDLTWSYA	444	
Db	361	GTYSSTSSSTYSIVDAVKTFADGPFVSIYETHAASNGSMSEQYDKSDGEQLAARDLTWSYA	420	
Qy	445	ALLTANNRRNSVVPASWGETSASSIPGICATSAIGTSSVTYVNSWSIYVATGGTTTAT	504	
Db	421	ALLTANNRRNSVVPASWGETSASSIPGICATSAIGTSSVTYVNSWSIYVATGGTTTAT	480	
Qy	505	PTGSGSVTSTSKTTATATSKTSTTTS	530	
Db	481	PTGSGSVTSTSKTTATATSKTSTTTS	506	
RESULT 14				
AAAM55976				
ID	AAAM55976	standard; protein; 616 AA.		
XX	AAAM55976;			
AC				
DT	27-JUL-1998	(first entry)		
XX				
DE	Aspergillus awamori	glucoamylase mutant N20C, A27C.		
XX				
KW	Aspergillus awamori; glucoamylase; Aspergillus sp; mutant; fungal; food;			
KW	fructose; corn; sweetener; 1,4-alpha-D-glucan glucohydrolase;			
KW	genetic engineering.			
XX				
OS	Synthetic.			
OS	Aspergillus awamori.			
XX				
FT	Key	Location/Qualifiers		
FT	Disulfide-bond	20..27		
XX				
PN	W09803639-A1.			
PD	29-JAN-1998.			
XX				
PF	24-JUL-1997;	97WO-US012983.		
XX				
PR	24-JUL-1996;	96US-0022578P.		
PR	02-AUG-1996;	96US-0023077P.		
XX				
PA	(IOWA ) UNIV IOWA STATE RES FOUND INC.			
XX				
PI	Allen M, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzatko R;			
PI	Ford C;			
DR	WPI; 1998-120764/11.			
XX				
PT	Genetically engineered fungal glucoamylase - useful in, e.g. food			
PT	industry for production of high fructose corn sweeteners.			
XX				
PS	Claim 1; Page; 97pp; English.			
XX				
CC	The present sequence represents a specifically claimed mutant			
CC	glucoamylase from Aspergillus awamori (1,4-alpha-D-glucan			
CC	glucohydrolase). The present invention describes fungal glucoamylases			
CC	(FG) comprising: a mutation pair Asn20Cys coupled with Ala27Cys forming a			
CC	disulfide bond between the 2 members of the pair; and a 311-314loop or			
CC	Seu41ALA mutation. FG can be used in industry for the production of high			
CC	fructose corn sweeteners, while the glucose produced by glucoamylase can			
CC	be crystallised or used in fermentation to produce organic products, e.g			
CC	citric acid, ascorbic acid, lysine, glutamic acid or ethanol for			
CC	beverages and fuel. The mutations provide increased thermal stability,			
CC	reduced isomaltose formation and increased pH optimum. N.B. The present			
CC	sequence is not given in the specification but is derived form SEQ ID			
CC	NO.1 as stated in the claim			
XX				
XX				
50	Sequence 616 AA;			

Best Local Similarity 99.2%; Pred. No. 7.3e-202;  
Matches 502; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
QY 25 ATLDSWLSNEATVARTAIINIGADGAWVSGADSGIVASPSTNDNDYFYTWTRDSGLYL 84
DB 1 ATLDSWLSNEATVARTAIINIGADGAWVSGADSGIVASPSTNDNDYFYTWTRDSGLYL 60
QY 85 KTLVDLFRNGDTSLSLSTIENYISAQAIYVGGISNPSGDLSSGAGLGPKFKNVDETAATGSM 144
DB 61 KTLVDLFRNGDTSLSLSTIENYISAQAIYVGGISNPSGDLSSGAGLGPKFKNVDETAATGSM 120
QY 145 GRPORDPALRATAMIGFGOMLLDNGYSTATDIWPLVRNDLSYVAQYWNQGYDLMEB 204
DB 121 GRPORDPALRATAMIGFGOMLLDNGYSTATDIWPLVRNDLSYVAQYWNQGYDLMEB 180
QY 205 VNGSSFFTIYAVOHRALVEGSAFATAVGSSCSMCDGAPILLCYLQSFMTGSFILLANFDS 264
DB 181 VNGSSFFTIYAVOHRALVEGSAFATAVGSSCSMCDGAPILLCYLQSFMTGSFILLANFDS 240
QY 265 RSGKDANTLLGSIHTFDPPEACDDSTFQPCSPRALANKEVDSFRSITTLNDGLSDSEA 324
DB 241 RSGKDANTLLGSIHTFDPPEACDDSTFQPCSPRALANKEVDSFRSITTLNDGLSDSEA 300
QY 325 VAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMDKQSLFVTDVSLDFPKALYSDAAT 384
DB 301 VAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMDKQSLFVTDVSLDFPKALYSDAAT 360
QY 385 GTYSSSSSTYSSIVDAVKTFADGFVSIETHAASNSMSEQYDKSGEQLSARDLTWSYA 444
DB 361 GTYSSSSSTYSSIVDAVKTFADGFVSIETHAASNSMSEQYDKSGEQLSARDLTWSYA 420
QY 445 ALLTANRRNSVVPASWGETSASVPGTCAATSAIGTYSVTYTSNPSIVATGCTTTTAT 504
DB 421 ALLTANRRNSVVPASWGETSASVPGTCAATSAIGTYSVTYTSNPSIVATGCTTTTAT 480
QY 505 PTGSGSVTSTSKTTATASKTSTTRS 530
DB 481 PTGSGSVTSTSKTTATASKTSTTRS 506
```

## RESULT 15

AAB15178 standard; protein; 616 AA.

AC AAB15178;

DT 11-DEC-2000 (first entry)

DE Aspergillus awamori mutant glucoamylase N20C/A27C substitution.

KM Glucoamylase; enzyme; carbohydrate; glucose;  
1,4-alpha-D-glucan glucohydrolase; mucin; mutation.

OS Aspergillus awamori.  
XX Synthetic.

XX Key Location/Qualifiers

FT Disulfide-bond 20..27

FT Misc-difference 20 /note= "Wild-type Asn substituted by Cys"

FT Misc-difference 27 /note= "Wild-type Ala substituted by Cys"

PN WO200043504-A1.

PD 27-JUL-2000.

PF 10-JAN-2000; 2000MO-US000532.

PR 22-JAN-1999; 99US-00236063.

PA (IOWA ) UNIV IOWA STATE RES FOUND INC.

XX

PI Allen MJ, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzacko R;  
PI Ford C;  
XX  
XX MPI; 2000-514725/46.

PT Fungal glucoamylase for selective production of glucose rather than alpha  
PT -1,6 linked disaccharide isomaltose, has mutation pair Asn20Cys coupled  
PT with Ala27Cys forming disulfide bond between the two stabilizing members.  
XX  
XX Claim 1; Page; 160pp; English.

XX Glucoamylase (1,4-alpha-D-glucan glucohydrolase; E.C. 3.2.1.3) is a  
CC carbohydrase. This enzyme cleaves D-glucose from the nonreducing ends of  
CC maltooligosaccharides, attacking alpha-(1,4)-, and alpha-(1,6)-glucosidic  
CC bonds. The present invention relates to mutant glucoamylases, which have  
CC increased thermostability, increased pH optimum and reduced isomaltose  
CC formation. The mutant proteins are useful for the selective production of  
CC glucose rather than alpha-1,6 linked disaccharide isomaltose. The present  
CC sequence is one such mutant enzyme. Note: The present sequence is not  
CC shown in the specification but is derived from the Aspergillus awamori  
CC wild-type glucoamylase sequence given in pages 152-153 of the sequence  
CC listing (SEQ ID 1)

XX Sequence 616 AA;

XX Query Match 94.4%; Score 2615; DB 3; Length 616;

XX Best Local Similarity 99.2%; Pred. No. 7.3e-202;  
XX Matches 502; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
QY 25 ATLDSWLSNEATVARTAIINIGADGAWVSGADSGIVASPSTNDNDYFYTWTRDSGLYL 84
DB 1 ATLDSWLSNEATVARTAIINIGADGAWVSGADSGIVASPSTNDNDYFYTWTRDSGLYL 60
QY 85 KTLVDLFRNGDTSLSLSTIENYISAQAIYVGGISNPSGDLSSGAGLGPKFKNVDETAATGSM 144
DB 61 KTLVDLFRNGDTSLSLSTIENYISAQAIYVGGISNPSGDLSSGAGLGPKFKNVDETAATGSM 120
QY 145 GRPORDPALRATAMIGFGOMLLDNGYSTATDIWPLVRNDLSYVAQYWNQGYDLMEB 204
DB 121 GRPORDPALRATAMIGFGOMLLDNGYSTATDIWPLVRNDLSYVAQYWNQGYDLMEB 180
QY 205 VNGSSFFTIYAVOHRALVEGSAFATAVGSSCSMCDGAPILLCYLQSFMTGSFILLANFDS 264
DB 181 VNGSSFFTIYAVOHRALVEGSAFATAVGSSCSMCDGAPILLCYLQSFMTGSFILLANFDS 240
QY 265 RSGKDANTLLGSIHTFDPPEACDDSTFQPCSPRALANKEVDSFRSITTLNDGLSDSEA 324
DB 241 RSGKDANTLLGSIHTFDPPEACDDSTFQPCSPRALANKEVDSFRSITTLNDGLSDSEA 300
QY 325 VAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMDKQSLFVTDVSLDFPKALYSDAAT 384
DB 301 VAVGRYPEDTYNNGNPMFLCTLAABOLYDALYQMDKQSLFVTDVSLDFPKALYSDAAT 360
QY 385 GTYSSSSSTYSSIVDAVKTFADGFVSIETHAASNSMSEQYDKSGEQLSARDLTWSYA 444
DB 361 GTYSSSSSTYSSIVDAVKTFADGFVSIETHAASNSMSEQYDKSGEQLSARDLTWSYA 420
QY 445 ALLTANRRNSVVPASWGETSASVPGTCAATSAIGTYSVTYTSNPSIVATGCTTTTAT 504
DB 421 ALLTANRRNSVVPASWGETSASVPGTCAATSAIGTYSVTYTSNPSIVATGCTTTTAT 480
QY 505 PTGSGSVTSTSKTTATASKTSTTRS 530
DB 481 PTGSGSVTSTSKTTATASKTSTTRS 506
```

## RESULT 16

AAB15184 standard; protein; 616 AA.

AC AAB15184;

DT 11-DEC-2000 (first entry)

```
XX A.awamori mutant glucoamylase S30P/E137A/S411A loop substitution.
DE
XX glucoamylase; enzyme; carboxylase; glucose;
KM 1,4-alpha-D-glucan glucosylhydrolase; mutuin; mutation.
XX
XX Aspergillus awamori.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH MISC-difference 30 /note= "Wild-type Ser substituted by Pro"
FT MISC-difference 137 /note= "Wild-type Glu substituted by Ala"
FT MISC-difference 411 /note= "Wild-type Ser substituted by Ala"
FT MISC-difference 411 /note= "Wild-type Ser substituted by Ala"
PN W0200043504-A1.
XX
XX 27-JUL-2000.
XX
XX 10-JAN-2000; 2000MO-US000532.
XX
XX 22-JAN-1999; 99US-00236063.
XX
XX (IOWA ) UNIV IOWA STATE RES FOUND INC.
XX
XX Allen MJ, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzatko R;
XX Ford C;
XX WPI; 2000-514725/46.
XX
XX Fungal glucoamylase for selective production of glucose rather than alpha
XX PT -1,6 linked disaccharide isomaltose, has mutation pair Asn20Cys coupled
XX PT with Ala27Cys forming disulfide bond between the two stabilizing members.
XX
XX
XX Claim 45; Page; 160pp; English.
XX
XX glucoamylase (1,4-alpha-D-glucan glucosylhydrolase; E.C. 3.2.1.3) is a
XX CC carboxylase. This enzyme cleaves D-glucose from the nonreducing ends of
XX CC maltooligosaccharides, attacking alpha-(1,4)-, and alpha-(1,6)-glucosidic
XX CC bonds. The present invention relates to mutant glucoamylases, which have
XX CC increased thermostability, increased pH optimum and reduced isomaltose
XX CC formation. The mutant proteins are useful for the selective production of
XX CC glucose rather than alpha-1,6 linked disaccharide isomaltose. The present
XX CC sequence is one such mutant enzyme. Note: The present sequence is not
XX CC shown in the specification but is derived from the Aspergillus awamori
XX CC wild-type glucoamylase sequence given in pages 152-153 of the sequence
XX CC listing (SEQ ID 1)
XX
XX Sequence 616 AA;
SQ
Query Match 94.3%; Score 2614; DB 3; Length 616;
Best Local Similarity 99.0%; Pred. No. 8.8e-202;
Matches 501; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 25 ATLDWSLNSNATYARATLANNIGADGAMVSGADSGIVVASPSTNDPDYFTWTRDSGLVL 84
DB 1 ATLDWSLNSNATYARATLANNIGADGAMVSGADSGIVVASPSTNDPDYFTWTRDSGLVL 60
QY 85 KTIYVDLFRNGDTSLSITENYISAOAIVOGISNPSGLSGAGIGEKFNVDETAYTGSW 144
DB 61 KTIYVDLFRNGDTSLSITENYISAOAIVOGISNPSGLSGAGIGEKFNVDETAYTGSW 120
QY 145 GRPORDPALRATMIGEGWLDNGYTSATDITVPLVNRNDLSYVAQYNNNGYDLME 204
DB 121 GRPORDPALRATMIGEGWLDNGYTSATDITVPLVNRNDLSYVAQYNNNGYDLME 180
QY 205 VNGSSFTTIAVORALVEGSAFATAVNGSSGWCDSQAPELLCYLQSFMTGSFILANPDS 264
DB 181 VNGSSFTTIAVORALVEGSAFATAVNGSSGWCDSQAPELLCYLQSFMTGSFILANPDS 240
QY 265 RSGKDANTLLGSIHTPEPAACDDSTFQPCSPRALANHKEVVDSPRSITTLNDGLSDSEA 324
```

```
DB 241 RSGKDANTLLGSIHTPEPAACDDSTFQPCSPRALANHKEVVDSPRSITTLNDGLSDSEA 300
QY 325 VAVGRYPEDTYVNGNFWPLCTTLAAEQLYDALYQMDKQSLLEVTVSLDFPKALYSDAAT 364
DB 301 VAVGRYPEDTYVNGNFWPLCTTLAAEQLYDALYQMDKQSLLEVTVSLDFPKALYSDAAT 360
QY 385 GTYSSSSSTYSSIVDAVKTFADGFVSIIVETHAASNSMSBQYDKSGEQLAARDLTWSYA 444
DB 361 GTYSSSSSTYSSIVDAVKTFADGFVSIIVETHAASNSMSBQYDKSGEQLAARDLTWSYA 420
QY 445 ALLTANNRRNSVVPASMGTSASVPGCAATSIGTYSVTYTSPMSIVATGTTTAT 504
DB 421 ALLTANNRRNSVVPASMGTSASVPGCAATSIGTYSVTYTSPMSIVATGTTTAT 460
QY 505 PTGSGSVTSKTKTATASKTSTTTS 530
DB 481 PTGSGSVTSKTKTATASKTSTTSS 506
RESULT 17
ID AAM55980
XX AAM55980 standard; protein; 616 AA.
XX
XX AAM55980;
XX
XX 27-JUL-1998 (first entry)
XX
XX Aspergillus awamori glucoamylase mutant N20C, A27C, S411A.
XX
XX Aspergillus awamori; glucoamylase; Aspergillus sp; mutant; fungal; food;
XX KM fructose; corn; sweetener; 1,4-alpha-D-glucan glucosylhydrolase;
XX KM genetic engineering.
XX
XX OS Synthetic.
XX OS Aspergillus awamori.
XX
XX Key Location/Qualifiers
FH Disulfide-bond 20..27
XX
XX W09803639-A1.
XX
XX 29-JAN-1998.
XX
XX 24-JUL-1997; 97MO-US012983.
XX
XX 24-JUL-1996; 96US-0022578P.
XX PR 02-AUG-1996; 96US-0023077P.
XX
XX (IOWA ) UNIV IOWA STATE RES FOUND INC.
XX
XX Allen M, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzatko R;
XX Ford C;
XX WPI; 1998-120764/11.
XX
XX Genetically engineered fungal glucoamylase - useful in, e.g. food
XX PT industry for production of high fructose corn sweeteners.
XX
XX Claim 14; Page; 97pp; English.
XX
XX The present sequence represents a specifically claimed mutant
XX CC glucoamylase from Aspergillus awamori (1,4-alpha-D-glucan
XX CC glucosylhydrolase). The present invention describes fungal glucoamylases
XX CC (FG) comprising: a mutation pair Asn20Cys coupled with Ala27Cys forming a
XX CC disulfide bond between the 2 members of the pair; and a 311-314loop or
XX CC Ser411Ala mutation. FG can be used in industry for the production of high
XX CC fructose corn sweeteners, while the glucose produced by glucoamylase can
XX CC be crystallized or used in fermentation to produce organic products, e.g.
XX CC citric acid, ascorbic acid, lysine, glutamic acid or ethanol for
XX CC beverages and fuel. The mutations provide increased thermal stability,
XX CC reduced isomaltose formation and increased pH optimum. N.B. The present
XX CC sequence is not given in the specification but is derived from SEQ ID
```

CC NO:1 as stated in the claim  
XX  
SQ Sequence 616 AA;

Query Match 94.3%; Score 2612; DB 2; Length 616;  
Best Local Similarity 99.0%; Pred. No. 1.3e-201;  
Matches 501; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

OY 25 ATLDSWLSNEATVARTAILNINIGADGAWYSGADSGIVVSPSTNDNDYFYTWTRDSGLYL 84
DB 1 ATLDSWLSNEATVARTAILNINIGADGAWYSGADSGIVVSPSTNDNDYFYTWTRDSGLYL 60
OY 85 KTLVDLFRNGDTSLSLTSTENIYSAQAIYVQISNPSGDLSSGAGLGPKFNVDEATYTGSM 144
DB 61 KTLVDLFRNGDTSLSLTSTENIYSAQAIYVQISNPSGDLSSGAGLGPKFNVDEATYTGSM 120
OY 145 GRPORDPALRATAMIGFCQWMLDNGYSTATDIWPLVRNDLSYVAQYWNQGTGYLWEE 204
DB 121 GRPORDPALRATAMIGFCQWMLDNGYSTATDIWPLVRNDLSYVAQYWNQGTGYLWEE 180
OY 205 VNGSSPFTIAYOHRALVEGSAFATAVGSSCWCDSQAPILCYLOSFWTGSFTLANFDS 264
DB 181 VNGSSPFTIAYOHRALVEGSAFATAVGSSCWCDSQAPILCYLOSFWTGSFTLANFDS 240
OY 265 RSGKDANTLLGSIHTFDPKACDSTFQPCSPRALANHEVDSFRSIYTLNDGLSDSEA 324
DB 241 RSGKDANTLLGSIHTFDPKACDSTFQPCSPRALANHEVDSFRSIYTLNDGLSDSEA 300
OY 325 VAVGRYPEDTYNGNFWPLCTLAAEQYDALYQMDKQSLVETDVSLEDFKALYSDDAT 384
DB 301 VAVGRYPEDTYNGNFWPLCTLAAEQYDALYQMDKQSLVETDVSLEDFKALYSDDAT 360
OY 385 GYSSSSSTYSSTIVDAVKTFADGFSIVETHAASNGSMSEQYDKSDGEQLAARDLTWSEA 444
DB 361 GYSSSSSTYSSTIVDAVKTFADGFSIVETHAASNGSMSEQYDKSDGEQLAARDLTWSEA 420
OY 445 ALLTANNRNSVVPASWGETSASVPGTCAATSAIGTYSVTYTSWPSIYATGCTTTAT 504
DB 421 ALLTANNRNSVVPASWGETSASVPGTCAATSAIGTYSVTYTSWPSIYATGCTTTAT 480
OY 505 PTGSGSVTSTSKTTATASKTSTTRS 530
DB 481 PTGSGSVTSTSKTTATASKTSTTRS 506

RESULT 18
AAB15181
ID AAB15181 standard; protein; 616 AA.
XX
XX AAB15181;
XX
DB 11-DEC-2000 (first entry)
XX
XX Aspergillus awamori mutant glucoamylase N20C/A27C/S411A substitution.
XX
XX Glucoamylase; enzyme; carbohydrate; glucose;
XX
XX 1,4-alpha-D-glucan glucohydrolase; muten; mutation.
XX
XX Aspergillus awamori.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX
XX Disulfide-bond 20. 27
XX
XX Misc-difference 20 /note= "Wild-type Asn substituted by Cys"
XX
XX Misc-difference 27 /note= "Wild-type Ala substituted by Cys"
XX
XX Misc-difference 411 /note= "Wild-type Ser substituted by Ala"
XX
XX WO200043504-A1.
XX
XX 27-JUL-2000.
XX
XX
```

XX 10-JAN-2000; 2000WO-US000532.  
XX  
XX 22-JAN-1999; 99US-00236063.  
XX  
XX (IOWA) UNIV IOWA STATE RES FOUND INC..  
XX  
XX Allen MJ, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzatko R;  
XX Ford C;  
XX  
XX MPI; 2000-514725/46.

PT Fungal glucoamylase for selective production of glucose rather than alpha  
PT -1,6 linked disaccharide isomaltose, has mutation pair Asn20Cys coupled  
PT with Ala27Cys forming disulfide bond between the two stabilizing members.  
XX  
XX  
XX Claim 14; Page; 160p; English.

XX Glucoamylase (1,4-alpha-D-glucan glucohydrolase; E.C. 3.2.1.3) is a  
CC carbohydrase. This enzyme cleaves D-glucose from the nonreducing ends of  
CC maltooligosaccharides, attacking alpha-(1,4)-, and alpha-(1,6)-glucosidic  
CC bonds. The present invention relates to mutant glucoamylases, which have  
CC increased thermostability, increased pH optimum and reduced isomaltose  
CC formation. The mutant proteins are useful for the selective production of  
CC glucose rather than alpha-1,6 linked disaccharide isomaltose. The present  
CC sequence is one such mutant enzyme. Note: The present sequence is not  
CC shown in the specification but is derived from the Aspergillus awamori  
CC wild-type glucoamylase sequence given in pages 152-153 of the sequence  
CC listing (SEQ ID 1)  
XX  
XX SQ Sequence 616 AA;

Query Match 94.3%; Score 2612; DB 3; Length 616;  
Best Local Similarity 99.0%; Pred. No. 1.3e-201;  
Matches 501; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

OY 25 ATLDSWLSNEATVARTAILNINIGADGAWYSGADSGIVVSPSTNDNDYFYTWTRDSGLYL 84
DB 1 ATLDSWLSNEATVARTAILNINIGADGAWYSGADSGIVVSPSTNDNDYFYTWTRDSGLYL 60
OY 85 KTLVDLFRNGDTSLSLTSTENIYSAQAIYVQISNPSGDLSSGAGLGPKFNVDEATYTGSM 144
DB 61 KTLVDLFRNGDTSLSLTSTENIYSAQAIYVQISNPSGDLSSGAGLGPKFNVDEATYTGSM 120
OY 145 GRPORDPALRATAMIGFCQWMLDNGYSTATDIWPLVRNDLSYVAQYWNQGTGYLWEE 204
DB 121 GRPORDPALRATAMIGFCQWMLDNGYSTATDIWPLVRNDLSYVAQYWNQGTGYLWEE 180
OY 205 VNGSSPFTIAYOHRALVEGSAFATAVGSSCWCDSQAPILCYLOSFWTGSFTLANFDS 264
DB 181 VNGSSPFTIAYOHRALVEGSAFATAVGSSCWCDSQAPILCYLOSFWTGSFTLANFDS 240
OY 265 RSGKDANTLLGSIHTFDPKACDSTFQPCSPRALANHEVDSFRSIYTLNDGLSDSEA 324
DB 241 RSGKDANTLLGSIHTFDPKACDSTFQPCSPRALANHEVDSFRSIYTLNDGLSDSEA 300
OY 325 VAVGRYPEDTYNGNFWPLCTLAAEQYDALYQMDKQSLVETDVSLEDFKALYSDDAT 384
DB 301 VAVGRYPEDTYNGNFWPLCTLAAEQYDALYQMDKQSLVETDVSLEDFKALYSDDAT 360
OY 385 GYSSSSSTYSSTIVDAVKTFADGFSIVETHAASNGSMSEQYDKSDGEQLAARDLTWSEA 444
DB 361 GYSSSSSTYSSTIVDAVKTFADGFSIVETHAASNGSMSEQYDKSDGEQLAARDLTWSEA 420
OY 445 ALLTANNRNSVVPASWGETSASVPGTCAATSAIGTYSVTYTSWPSIYATGCTTTAT 504
DB 421 ALLTANNRNSVVPASWGETSASVPGTCAATSAIGTYSVTYTSWPSIYATGCTTTAT 480
OY 505 PTGSGSVTSTSKTTATASKTSTTRS 530
DB 481 PTGSGSVTSTSKTTATASKTSTTRS 506
```



```

RESULT 19
AAMS5977
ID AAM55977 standard; protein; 616 AA.
XX
AC AAM55977;
XX
DT 27-JUL-1998 (first entry)
XX
DE Aspergillus awamori glucoamylase mutant N20C, A27C, S30P, G137A.
XX
KM Aspergillus awamori; glucoamylase; Aspergillus sp; mutant; fungal; food;
KM fructose; corn; sweetener; 1,4-alpha-D-glucan glucosyltransferase;
KM genetic engineering.
XX
OS Synthetic.
XX
OS Aspergillus awamori.
XX
FH Key Location/Qualifiers
FH Disulfide-bond 20..27
XX
PN WO9803639-A1.
XX
PD 29-JAN-1998.
XX
PF 24-JUL-1997; 97WO-US012983.
XX
PR 24-JUL-1996; 96US-0022578P.
PR 02-AUG-1996; 96US-0023077P.
XX
XX
PA (IOWA ) UNIV IOWA STATE RES FOUND INC.
XX
PI Allen M, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzatko R;
PI Ford C;
XX
DR WPI; 1998-120764/11.
XX
PT Genetically engineered fungal glucoamylase - useful in, e.g. food
PT industry for production of high fructose corn sweeteners.
XX
PS Claim 4; Page; 97P; English.
XX
XX
CC The present sequence represents a specifically claimed mutant
CC glucoamylase from Aspergillus awamori (1,4-alpha-D-glucan
CC glucosyltransferase). The present invention describes fungal glucoamylases
CC (FG) comprising: a mutation pair Asn20Cys coupled with Ala27Cys forming a
CC disulphide bond between the 2 members of the pair; and a 311-314loop or
CC Ser411Ala mutation. FG can be used in industry for the production of high
CC fructose corn sweeteners, while the glucose produced by glucoamylase can
CC be crystallised or used in fermentation to produce organic products, e.g.
CC citric acid, ascorbic acid, lysine, glutamic acid or ethanol for
CC beverages and fuel. The mutations provide increased thermal stability,
CC reduced isomaltose formation and increased pH optimum. N.B. The present
CC sequence is not given in the specification but is derived from SEQ ID
CC NO:1 as stated in the claim
XX
XX
SQ Sequence 616 AA;
XX
XX
Query Match 94.0%; Score 2604; DB 2; Length 616;
Best Local Similarity 98.8%; Pred. No. 5.6e-201;
Matches 500; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
XX
QY 25 ATLDWSLSEATVARTAIILNIGADGAVSGADSGIVVSPSTNDPDYFTYTRDSGLVL 84
DB 1 ATLDWSLSEATVARTAIILNIGADGCMVPGADSGIVVSPSTNDPDYFTYTRDSGLVL 60
XX
QY KTIIVDLFRNDGTSLSLTENYISAQAVOGISNPSGSLSGAGIGEPKENVDEATAYGWSV 144
DB 61 KTIIVDLFRNDGTSLSLTENYISAQAVOGISNPSGSLSGAGIGEPKENVDEATAYGWSV 120
XX
QY 145 GRPORDPALRATAMIGFGOMLDNGYSTATDITVWPLVLRNDLSYVAQYNNQGYDLMEZ 204
DB 121 GRPORDPALRATAMIGFGOMLDNGYSTATDITVWPLVLRNDLSYVAQYNNQGYDLMEZ 180
XX

```

```

QY 205 VNGSFFTTAVQHRALVBSAPATAVGSSCSWCDQAPILCYLQSFMTGSFLLANFDS 264
DB 181 VNGSFFTTAVQHRALVBSAPATAVGSSCSWCDQAPILCYLQSFMTGSFLLANFDS 240
XX
QY 265 RSGKDANTLLGSITHTDPEAACDSTFOPCSPRALNHEVVDSPFSIYTLINDGLSDSEA 324
DB 241 RSGKDANTLLGSITHTDPEAACDSTFOPCSPRALNHEVVDSPFSIYTLINDGLSDSEA 300
XX
QY 325 VAVGRYPEDTYNGNPMFLCTLAABEQLYDALYQMDKQSLSEVTVSLDFPKALYSDDAT 384
DB 301 VAVGRYPEDTYNGNPMFLCTLAABEQLYDALYQMDKQSLSEVTVSLDFPKALYSDDAT 360
XX
QY 385 GTYSSSSSTYSSIVDAVKTPADGPFVSIIVETHAASNGSMSEQYDKSDGEOLASADLTWSYA 444
DB 361 GTYSSSSSTYSSIVDAVKTPADGPFVSIIVETHAASNGSMSEQYDKSDGEOLASADLTWSYA 420
XX
QY 445 ALLTANNRRNSVVPAPMGETSASVPGTCAATSAIGTSSVTVTSPSIVATGCTTTTAT 504
DB 421 ALLTANNRRNSVVPAPMGETSASVPGTCAATSAIGTSSVTVTSPSIVATGCTTTTAT 480
XX
QY 505 PTGSGSVTSTSKTTATATASKTSTTRS 530
DB 481 PTGSGSVTSTSKTTATATASKTSTTRS 506
XX

```

```

RESULT 20
AAB15179
ID AAB15179 standard; protein; 616 AA.
XX
AC AAB15179;
XX
DT 11-DEC-2000 (first entry)
XX
DE A. awamori mutant glucoamylase N20C/A27C/S30P/G137A substitution.
XX
KM Glucoamylase; enzyme; carbohydrase; glucose;
KM 1,4-alpha-D-glucan glucosyltransferase; mutain; mutation.
XX
OS Aspergillus awamori.
OS Synthetic.
XX
FH Key Location/Qualifiers.
FH Disulfide-bond 20..27
FT Misc-difference 20 /note= "wild-type Asn substituted by Cys"
FT FT Misc-difference 27 /note= "wild-type Ala substituted by Cys"
FT FT Misc-difference 30 /note= "wild-type Ser substituted by Pro"
FT FT Misc-difference 137 /note= "wild-type Gly substituted by Ala"
XX
PN WO200043504-A1.
XX
PD 27-JUL-2000.
XX
PF 10-JAN-2000; 2000WO-US000532.
XX
PR 22-JAN-1999; 99US-00236063.
XX
XX
PA (IOWA ) UNIV IOWA STATE RES FOUND INC.
XX
PI Allen M, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzatko R;
PI Ford C;
XX
DR WPI; 2000-514725/46.
XX
XX
PT Fungal glucoamylase for selective production of glucose rather than alpha
PT -1,6 linked disaccharide isomaltose, has mutation pair Asn20Cys coupled
PT with Ala27Cys forming disulfide bond between the two stabilizing members.
XX
PS Claim 4; Page; 160pp; English.
XX

```

CC Glucoamylase (1,4-alpha-D-glucan glucosyltransferase; E.C. 3.2.1.3) is a  
 CC carbohydrase. This enzyme cleaves D-glucose from the nonreducing ends of  
 CC maltooligosaccharides, attacking alpha-(1,4)- and alpha-(1,6)-glucosidic  
 CC bonds. The present invention relates to mutant glucoamylases, which have  
 CC increased thermostability, increased pH optimum and reduced isomaltose  
 CC formation. The mutant proteins are useful for the selective production of  
 CC glucose rather than alpha-1,6 linked disaccharide isomaltose. The present  
 CC sequence is one such mutant enzyme. Note: The present sequence is not  
 CC shown in the specification but is derived from the Aspergillus awamori  
 CC wild-type glucoamylase sequence given in pages 152-153 of the sequence  
 CC listing (SEQ ID 1)

XX Sequence 616 AA:

Query Match 94.0%; Score 2604; DB 3; Length 616;  
 Best Local Similarity 98.8%; Pred. No. 5.6e-201;  
 Matches 500; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 25 ATLDLWLSNEATVARTALINNIAGDAGAWVSGADSGIVVSPSTNDPDYFTWTRDSGLV 84  
 DB 1 ATLDLWLSNEATVARTALINNIAGDAGAWVSGADSGIVVSPSTNDPDYFTWTRDSGLV 60  
 QY 85 KTLVDLFRNGDTSLSLTIENTYISAQAIYVQGISNPSGDLSSGAGLGEKFNVDETAYTGSW 144  
 DB 61 KTLVDLFRNGDTSLSLTIENTYISAQAIYVQGISNPSGDLSSGAGLGEKFNVDETAYTGSW 120  
 QY 145 GRPORDPALRATAMIFGQWMLDNGYSTATDIWPLVRNDLSYVAQVWNGYDLMEE 204  
 DB 121 GRPORDPALRATAMIFGQWMLDNGYSTATDIWPLVRNDLSYVAQVWNGYDLMEE 180  
 QY 205 VNGSSFFTTAVOHRALVEGSAFATVAVGSSCSWCDQAPILCYLQSFMTGSFTLANPDS 264  
 DB 181 VNGSSFFTTAVOHRALVEGSAFATVAVGSSCSWCDQAPILCYLQSFMTGSFTLANPDS 240  
 QY 265 RSGKANTLLGSIHTDPPEACDSTFQPCSPALANHEKVDSPRSITTLNDGSDSEA 324  
 DB 241 RSGKANTLLGSIHTDPPEACDSTFQPCSPALANHEKVDSPRSITTLNDGSDSEA 300  
 QY 325 VAVGRYPEDTYVNGNPMFLCTLAABOLYDALYOMKQSLVTVDSLPDFKALYSDAAT 384  
 DB 301 VAVGRYPEDTYVNGNPMFLCTLAABOLYDALYOMKQSLVTVDSLPDFKALYSDAAT 360  
 QY 385 GTYSSSSSTYSIVDAVKTFADGFVSIIVETHAASNGSMSEQYKSDGEOLASARDLWMSYA 444  
 DB 361 GTYSSSSSTYSIVDAVKTFADGFVSIIVETHAASNGSMSEQYKSDGEOLASARDLWMSYA 420  
 QY 445 ALLTANNRNSVVPASWGETSASSVPETCAATSAIGTYSSTVTSWPSIVATGCTTTAT 504  
 DB 421 ALLTANNRNSVVPASWGETSASSVPETCAATSAIGTYSSTVTSWPSIVATGCTTTAT 480  
 QY 505 PTGSGSVTSTSKTTATASKTSTTTS 530  
 DB 481 PTGSGSVTSTSKTTATASKTSTTTS 506

RESULT 21

AAW55978  
 ID AAW55978 standard; protein; 621 AA.

AC AAW55978;  
 XX 27-JUL-1998 (first entry)

DE Aspergillus awamori glucoamylase mutant 311-314 loop mutation.

XX Aspergillus awamori; glucoamylase; Aspergillus sp; mutant; fungal; food;  
 KW fructose; corn; sweetener; 1,4-alpha-D-glucan glucosyltransferase;  
 XX genetic engineering.

OS Synthetic.  
 OS Aspergillus awamori.  
 XX  
 FH Key Location/Qualifiers

PT Misc-difference 311..319  
 FT /notes "311-314 loop mutation"

XX MO9803639-A1.

XX 29-JAN-1998.

XX 24-JUL-1997; 97WO-US012983.

XX 24-JUL-1996; 96US-0022578P.

XX 02-AUG-1996; 96US-0023077P.

XX (IOWA ) UNIV IOWA STATE RES FOUND INC.

XX Allen M, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzakto R;  
 PI Ford C;  
 XX WPI; 1998-120764/11.

XX Genetically engineered fungal glucoamylase - useful in, e.g. food  
 PT industry for production of high fructose corn sweeteners.

XX Claim 7; Page; 97pp; English.

XX The present sequence represents a specifically claimed mutant  
 CC glucoamylase from Aspergillus awamori (1,4-alpha-D-glucan  
 CC glucosyltransferase). The present invention describes fungal glucoamylases  
 CC (FG) comprising: a mutation pair Asn20Cys coupled with Ala27Cys forming a  
 CC disulphide bond between the 2 members of the pair; and a 311-314loop or  
 CC Ser41Ala mutation. FG can be used in industry for the production of high  
 CC fructose corn sweeteners, while the glucose produced by glucoamylase can  
 CC be crystallised or used in fermentation to produce organic products, e.g.  
 CC citric acid, ascorbic acid, lysine, glutamic acid or ethanol for  
 CC beverages and fuel. The mutations provide increased thermal stability,  
 CC reduced isomaltose formation and increased pH optimum. N.B. The present  
 CC sequence is not given in the specification but is derived from SEQ ID  
 CC NO:1 as stated in the claim

XX Sequence 621 AA:

Query Match 93.8%; Score 2600.5; DB 2; Length 621;  
 Best Local Similarity 98.2%; Pred. No. 1.1e-200;  
 Matches 502; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

QY 25 ATLDLWLSNEATVARTALINNIAGDAGAWVSGADSGIVVSPSTNDPDYFTWTRDSGLV 84  
 DB 1 ATLDLWLSNEATVARTALINNIAGDAGAWVSGADSGIVVSPSTNDPDYFTWTRDSGLV 60  
 QY 85 KTLVDLFRNGDTSLSLTIENTYISAQAIYVQGISNPSGDLSSGAGLGEKFNVDETAYTGSW 144  
 DB 61 KTLVDLFRNGDTSLSLTIENTYISAQAIYVQGISNPSGDLSSGAGLGEKFNVDETAYTGSW 120  
 QY 145 GRPORDPALRATAMIFGQWMLDNGYSTATDIWPLVRNDLSYVAQVWNGYDLMEE 204  
 DB 121 GRPORDPALRATAMIFGQWMLDNGYSTATDIWPLVRNDLSYVAQVWNGYDLMEE 180  
 QY 205 VNGSSFFTTAVOHRALVEGSAFATVAVGSSCSWCDQAPILCYLQSFMTGSFTLANPDS 264  
 DB 181 VNGSSFFTTAVOHRALVEGSAFATVAVGSSCSWCDQAPILCYLQSFMTGSFTLANPDS 240  
 QY 265 RSGKANTLLGSIHTDPPEACDSTFQPCSPALANHEKVDSPRSITTLNDGSDSEA 324  
 DB 241 RSGKANTLLGSIHTDPPEACDSTFQPCSPALANHEKVDSPRSITTLNDGSDSEA 300  
 QY 325 VAVGRYPEDTYVNGNPMFLCTLAABOLYDALYOMKQSLVTVDSLPDFKALY 379  
 DB 301 VAVGRYPEDTYVNGNPMFLCTLAABOLYDALYOMKQSLVTVDSLPDFKALY 360  
 QY 380 SDAATGTYSSSSSTYSIVDAVKTFADGFVSIIVETHAASNGSMSEQYKSDGEOLASARDL 439  
 DB 361 SDAATGTYSSSSSTYSIVDAVKTFADGFVSIIVETHAASNGSMSEQYKSDGEOLASARDL 420  
 QY 440 TWSYALLTANNRNSVVPASWGETSASSVPETCAATSAIGTYSSTVTSWPSIVATGCT 499

```
Db 421 TWISYAAALITANNRNNSVVPASWGETSASVPGTCAATSAIGTYSVTVTWSPISVATGCT 480
Qy 500 TTTATPTGSGSVTSTSKTTATATASKTSTTTRS 530
Db 481 TTTATPTGSGSVTSTSKTTATATASKTSTTSS 511

RESULT 22
AAB15183 standard; protein; 621 AA.
XX
AC AAB15183;
XX
DT 11-DEC-2000 (first entry)
XX
DE A. awamori mutant glucoamylase S30P/G137A/311-314 loop substitution.
XX
KM Glucoamylase; enzyme; carbohydrase; glucose;
XX 1,4-alpha-D-glucan glucosyltransferase; mutacin; mutation.
XX
OS Aspergillus awamori.
XX Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 30 /note= "Wild-type Ser substituted by Pro"
FT Misc-difference 137 /note= "Wild-type Gly substituted by Ala"
FT Misc-difference 311..314 /note= "Wild-type Tyr-Tyr-Asn-Gly substituted by Tyr-Asn-Gly-Asn-Gly-Asn-Ser-Gln-Gly"
FT FT
XX
XX WO200043504-A1.
XX
XX 27-JUL-2000.
XX
XX 10-JAN-2000; 2000WO-US000532.
XX
XX 22-JAN-1999; 99US-00236063.
XX
XX (IOWA ) UNIV IOWA STATE RES FOUND INC.
XX
XX Allen MJ, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzalko R;
XX Ford C;
XX
XX WPI; 2000-514725/46.
XX
XX Fungal glucoamylase for selective production of glucose rather than alpha
XX 1,6 linked disaccharide isomaltose, has mutation pair Asn20Cys coupled
XX with Ala27Cys forming disulfide bond between the two stabilizing members.
XX
XX Claim 42; Page; 160pp; English.
XX
XX Glucoamylase (1,4-alpha-D-glucan glucosyltransferase; E.C. 3.2.1.3) is a
XX carbohydrase. This enzyme cleaves D-glucose from the nonreducing ends of
XX maltooligosaccharides, attacking alpha-(1,4)-, and alpha-(1,6)-glucosidic
XX bonds. The present invention relates to mutant glucoamylases, which have
XX increased thermostability, increased pH optimum and reduced isomaltose
XX formation. The mutant proteins are useful for the selective production of
XX glucose rather than alpha-1,6 linked disaccharide isomaltose. The present
XX sequence is one such mutant enzyme. Note: The present sequence is not
XX shown in the specification but is derived from the Aspergillus awamori
XX wild-type glucoamylase sequence given in pages 152-153 of the sequence
XX listing (SEQ ID 1)
XX
XX Sequence 621 AA;
XX
XX Query Match 93.5%; Score 2589.5; DB 3; Length 621;
XX Best Local Similarity 97.8%; Pred. No. 8.5e-200;
XX Matches 500; Conservative 1; Mismatches 5; Indels 5; Gaps 1;
XX
XX 25 ATLDWSLSEATVARTAILANNIGADGAMVSGADGIIVASBSTDNPDTFYTTWRDGLVL 84
```

```
Db 1 ATLDWSLSEATVARTAILANNIGADGAMVPGADSGIVVAPSTDNPDYFTTWTRDGLVL 60
Qy 85 KTLVDFPRNGDTSILSTIENYISAQAIVOGINSPSGDLSSGAGLGEPKFVNDEATATGSAW 144
Db 61 KTLVDFPRNGDTSILSTIENYISAQAIVOGINSPSGDLSSGAGLGEPKFVNDEATATGSAW 120
Qy 145 GRPQRDGPALRATAMTIGFGQMLDNGYSTATDITVPLVRNDLSYVAQYWNQGTGYLMEW 204
Db 121 GRPQRDGPALRATAMTAFQGMMLDNGYSTATDITVPLVRNDLSYVAQYWNQGTGYLMEW 180
Qy 205 VNGSSEFTTAVQHRALVEGSAFATAVSSGSCWCDQAPETLCYLOSFWTGSFTLANPDS 264
Db 181 VNGSSEFTTAVQHRALVEGSAFATAVSSGSCWCDQAPETLCYLOSFWTGSFTLANPDS 240
Qy 265 RSGKDANTLLGSIHTDPPEAACDSTFQPCSPALANHKVVDSPFSIYTLNDGLSDSEA 324
Db 241 RSGKDANTLLGSIHTDPPEAACDSTFQPCSPALANHKVVDSPFSIYTLNDGLSDSEA 300
Qy 325 VAVGRYPEDTY-----YNGNPWFLLCTLAABEQLYDALYQMDKQSLVETDVSIDFFKALY 379
Db 301 VAVGRYPEDTYNNGNGNSQGNPWFLLCTLAABEQLYDALYQMDKQSLVETDVSIDFFKALY 360
Qy 380 SDAATGTYSSSSSTYSIVDAVKTFFADGFYSIVETHAASGWSBQYDKSDGQLSARDL 439
Db 361 SDAATGTYSSSSSTYSIVDAVKTFFADGFYSIVETHAASGWSBQYDKSDGQLSARDL 420
Qy 440 TWISYAAALITANNRNNSVVPASWGETSASVPGTCAATSAIGTYSVTVTWSPISVATGCT 499
Db 421 TWISYAAALITANNRNNSVVPASWGETSASVPGTCAATSAIGTYSVTVTWSPISVATGCT 480
Qy 500 TTTATPTGSGSVTSTSKTTATATASKTSTTTRS 530
Db 481 TTTATPTGSGSVTSTSKTTATATASKTSTTSS 511

RESULT 23
AAB15182 standard; protein; 621 AA.
XX
XX AAB15182;
XX
XX 11-DEC-2000 (first entry)
XX
XX A. awamori mutant glucoamylase N20C/A27C/S411A/311-314 loop substitution.
XX
XX Glucoamylase; enzyme; carbohydrase; glucose;
XX 1,4-alpha-D-glucan glucosyltransferase; mutacin; mutation.
XX
XX Aspergillus awamori.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Disulfide-bond 20..27
XX Misc-difference 20 /note= "Wild-type Asn substituted by Cys"
XX Misc-difference 27 /note= "Wild-type Ala substituted by Cys"
XX Misc-difference 311..314 /note= "Wild-type Tyr-Tyr-Asn-Gly substituted by Tyr-Asn-Gly-Asn-Gly-Asn-Ser-Gln-Gly"
XX Misc-difference 411 /note= "Wild-type Ser substituted by Ala"
XX
XX WO200043504-A1.
XX
XX 27-JUL-2000.
XX
XX 10-JAN-2000; 2000WO-US000532.
XX
XX 22-JAN-1999; 99US-00236063.
XX
XX (IOWA ) UNIV IOWA STATE RES FOUND INC.
XX
```

XX Allen MJ, Fang T, Li Y, Liu H, Chen H, Coutinho P, Honzacko R;  
PI Ford C;  
XX  
XX WPI: 2000-514725/46.  
XX  
XX Fungal glucanase for selective production of glucose rather than alpha  
PT -1,6 linked disaccharide isomaltose, has mutation pair Asn20Cys coupled  
PT with Ala27Cys forming disulfide bond between the two stabilizing members.  
XX  
XX Claim 16; Page; 160pp; English.  
XX  
XX Glucoamylase (1,4-alpha-D-glucan glucosylhydrolase; E.C. 3.2.1.3) is a  
CC carbohydrase. This enzyme cleaves D-glucose from the nonreducing ends of  
CC maltooligosaccharides, attacking alpha-(1,4)-, and alpha-(1,6)-glucosidic  
CC bonds. The present invention relates to mutant glucanases, which have  
CC increased thermostability, increased pH optimum and reduced isomaltose  
CC formation. The mutant proteins are useful for the selective production of  
CC glucose rather than alpha-1,6 linked disaccharide isomaltose. The present  
CC sequence is one such mutant enzyme. Note: The present sequence is not  
CC shown in the specification but is derived from the Aspergillus awamori  
CC wild-type glucanase sequence given in pages 152-153 of the sequence  
CC listing (SEQ ID 1)  
XX  
XX Sequence 621 AA;  
SQ

Query Match 93.3%; Score 2584.5; DB 3; Length 621;  
Best Local Similarity 97.7%; Pred. No. 2.1e-199;  
Matches 499; Conservative 2; Mismatches 5; Indels 5; Gaps 1;

QY 25 ATLDWSLNEATVARTAILNINIGADGAWVSGADGIVVSPSTNDPDEYTTWRDGLV 84  
DB 1 ATLDWSLNEATVARTAILNINIGADGAWVSGADGIVVSPSTNDPDEYTTWRDGLV 60

QY 85 KTLVDFRNGDTSLSLTENYISAQAIIVOGISNPSGDLSSGAGLGEPKENVDEATVATGSM 144  
DB 61 KTLVDFRNGDTSLSLTENYISAQAIIVOGISNPSGDLSSGAGLGEPKENVDEATVATGSM 120

QY 145 GRPDRGPRALRATAMIGFGOMLDNGYATATDVIWPLVANDLSYVAQVWNOGYLMEB 204  
DB 121 GRPDRGPRALRATAMIGFGOMLDNGYATATDVIWPLVANDLSYVAQVWNOGYLMEB 180

QY 205 VNGSFFFTIAVORALVEGSAFATAVGSSCSWCDQAPELLCYLQSFMTGSFTLANFDS 264  
DB 181 VNGSFFFTIAVORALVEGSAFATAVGSSCSWCDQAPELLCYLQSFMTGSFTLANFDS 240

QY 265 RSGKDANTLLGSIHTFDPEAACDSTFQPCSPALANHKEVDSFRSIYTLNDGLSDSEA 324  
DB 241 RSGKDANTLLGSIHTFDPEAACDSTFQPCSPALANHKEVDSFRSIYTLNDGLSDSEA 300

QY 325 VAVGRPEEDTY-----YNGNPWFLCTLAABEQYDALYQWDKGSLEVTDVSLDFPKALY 379  
DB 301 VAVGRPEEDTYNGNSQGNPWLCTLAABEQYDALYQWDKGSLEVTDVSLDFPKALY 360

QY 380 SDAATGTYSSSSSTYSIVDAVTFADGFVSIYETHAASNGSMEQYDSDGQOLARDU 439  
DB 361 SDAATGTYSSSSSTYSIVDAVTFADGFVSIYETHAASNGSMEQYDSDGQOLARDU 420

QY 440 TMSYALLTANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSVTVTSMPSIVATGGT 499  
DB 421 TMSYALLTANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSVTVTSMPSIVATGGT 480

QY 500 TTTATPTGGSGVTSSTKTTATAKTSTTTRS 530  
DB 481 TTTATPTGGSGVTSSTKTTATAKTSTTTRS 511

RESULT 24  
ABP96630  
ID ABP96630 standard; protein; 1095 AA.  
XX  
XX AC ABP96630;  
XX

DT 02-JUN-2003 (first entry)  
XX  
XX Alpha-amylase/glucanase fusion protein sequence SEQ ID NO:45.  
DE  
XX  
XX Self-processing plant; plant; processing enzyme; alpha-amylase; grain;  
XX pullulanase; alpha-glucosidase; glucose isomerase; glucanase;  
XX mesophilic; thermophilic; hyperthermophilic; transgenic plant; starch;  
XX maltodextrin; ethanol; fermentation; beverage; enzyme.  
OS  
OS Aspergillus shirousami.  
OS Synthetic.  
XX  
XX WO2003018766-A2.  
XX  
XX 06-MAR-2003.  
XX  
XX 27-AUG-2002; 2002WO-US027129.  
XX  
XX 27-AUG-2001; 2001US-0315281P.  
XX  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
XX Lanahan MB, Basu SS, Batie CJ, Chen W, Craig J, Kinkema M;  
PI  
XX  
XX WPI: 2003-268420/26.  
XX  
XX N-PSDB; ACC44572.  
XX  
XX Novel polynucleotide encoding hyperthermophilic processing enzymes e.g.  
PT alpha-amylase, useful for producing plant to produce food products having  
PT improved taste or fermentable substrates for ethanol.  
XX  
XX Claim 1; Page 107; 158pp; English.  
XX  
XX The present invention describes polynucleotides which encode processing  
CC enzymes (e.g. alpha-amylase, pullulanase, alpha-glucosidase, glucose  
CC isomerase, or glucanase) that are optimised for expression in plants.  
CC The polynucleotides encode mesophilic, thermophilic or hyperthermophilic  
CC processing enzymes, which are activated under suitable conditions to act  
CC upon the desired substrate. Also described are self-processing transgenic  
CC plants and plant parts, e.g. grain, which express one or more of these  
CC enzymes and have an altered composition that facilitates plant and grain  
CC processing. Also described is a method (M) for converting starch to  
CC starch-derived products in a transformed plant part (TRP), by activating  
CC the starch processing enzyme contained in it. Transgenic grain is useful  
CC for preparing maltodextrin. A transformed plant (TP) can be used to  
CC produce food products having improved taste and to produce fermentable  
CC substrates for ethanol and fermented beverages. (M) eliminates the need  
CC to mill or physically disrupt the integrity of plant parts prior to  
CC recovery of starch-derived products. The present sequence represents  
CC alpha-amylase/glucanase fusion protein, which is given in the  
CC exemplification of the present invention  
XX  
XX Sequence 1095 AA;  
SQ

Query Match 91.1%; Score 2523; DB 6; Length 1095;  
Best Local Similarity 92.9%; Pred. No. 4.7e-194;  
Matches 461; Conservative 16; Mismatches 15; Indels 6; Gaps 2;

QY 13 VCTGLANVISKATIDWSLNEATVARTAILNINIGADGAWVSGADGIVVSPSTNDPXY 72  
DB 474 ICSS-----SKPATIDWSLNEATVARTAILNINIGADGAWVSGADGIVVSPSTNDPXY 528

QY 73 FTTWRDGLVLTVDLFRNGDTSLSLTENYISAQAIIVOGISNPSGDLSSGAGLGEPK 132  
DB 529 FTTWRDGLVLTVDLFRNGDTSLSLTENYISAQAIIVOGISNPSGDLSSGAGLGEPK 587

QY 133 FNVDEATYGSWGRPORQSPALRATAMIGFGOMLDNGYATATDVIWPLVANDLSYVAQ 192  
DB 588 FNVDEATYGSWGRPORQSPALRATAMIGFGOMLDNGYATATDVIWPLVANDLSYVAQ 647

QY 193 YNNGTGYDIMEVNGSSFFTTIAVORALVEGSAFATAVGSSCSWCDQAPELLCYLQSF 252  
DB 648 YNNGTGYDIMEVNGSSFFTTIAVORALVEGSAFATAVGSSCSWCDQAPELLCYLQSF 707

QY 253 TGSFIIANFDSRSSGKDANTLLGSIHTFDEPAACDDSTFQPCSPRALANHKEVVDSPRSI 312  
| : : : : :  
Db 708 TGSYIIANFDSRSSGKDTNLLGSIHTFDEPACDDSTFQPCSPRALANHKEVVDSPRSI 767  
QY 313 YTLNDGSDSEAVANGVPEDTYNGNPMWFLCTLAABEQLYDALYQMDKQSGLEIVDVSL 372  
| : : : : :  
Db 768 YTLNDGSDSEAVANGVPEDSYNGNPMWFLCTLAABEQLYDALYQMDKQSGLEIVDVSL 827  
QY 373 DFFKALYSDAATGYSSSSSTYSIYDAVKTFADGFVSIYETHAASNGSMSEQYDKSGDGE 432  
| : : : : :  
Db 828 DFFKALYSGAATGYSSSSSTYSIYSAVKTFADGFVSIYETHAASNGSLSEQEDKSGDGD 887  
QY 433 QLSARDLTWSYAAALLTANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSVTWSPS 492  
| : : : : :  
Db 888 ELSARDLTWSYAAALLTANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSVTWSPS 947  
QY 493 IVATGCTTTATPTGSGSVTSTSKTTATASKTSTTTS 530  
| : : : : :  
Db 948 IVATGCTTTATPTGSGSVTSTSKTTATASKTSTTTS 985

## RESULT 25

ABM00045  
ID ABM00045 standard; protein; 471 AA.

XX AC ABM00045;

XX DT 02-APR-2003 (first entry)

XX DE AMG SEQ ID NO 36.

XX KM allergen; protein coordinate data; vaccine; anti-allergic; immunogenicity;  
XX KM detergent; personal care composition; cosmetic.

XX OS Unidentified.

XX PN WO200183559-A2.

XX PD 08-NOV-2001.

XX PF 30-APR-2001; 2001WO-DK000293.

XX PR 28-APR-2000; 2000DK-00000707.

XX PR 10-MAY-2000; 2000US-0203345P.

XX PR 28-FEB-2001; 2001DK-00000327.

XX PR 21-MAR-2001; 2001US-0277817P.

XX PA (NOVO ) NOVOZYMES AS.

XX PI Roggen EL, Ernst S, Svendsen A, Friis EP, Von Der Osten C;

XX DR WPI; 2001-626552/72.

XX PT Selecting protein variants having modified immunogenicity, used to  
XX PT produce vaccines, detergents and personal care compositions, involves  
XX PT localizing epitope sequences on the three-dimensional structure of a  
XX PT protein.

XX PS Claim 99; Page 510-512; 513pp; English.

XX CC The invention relates to selecting a protein variant having modified  
XX CC immunogenicity, compared to a parent protein, comprising using the  
XX CC antibody binding sequence to localise epitope sequences on the three  
XX CC dimensional structure of the parent protein and defining an epitope area  
XX CC including amino acids within 5 Angstrom of the epitope amino acids. The  
XX CC method is useful for identifying structural epitopes on the 3-dimensional  
XX CC surface of commercial and environmental allergens. Compositions  
XX CC containing the protein variants are used as vaccines, detergents and  
XX CC personal care compositions, e.g. shampoo, balsam, hair conditioners, hair  
XX CC cream, hair rinse, hair spray, chewing gum, skin cream, sunscreen,  
XX CC shaving foam, cream soap, skin milk or foundation. The present sequence

CC is that of a polypeptide of the invention  
XX Sequence 471 AA;

Query Match 89.1%; Score 2468; DB 4; Length 471;  
Best Local Similarity 100.0%; Pred. No. 3.5e-190; Indels 0; Gaps 0;  
Matches 471; Conservative 0; Mismatches 0;

QY 25 ATLDWSLNEATVARTAIINNIGADGAWVSGADSGIVVAS PSTNDPDEFTWTRDSGLVL 84  
| : : : : :  
Db 1 ATLDWSLNEATVARTAIINNIGADGAWVSGADSGIVVAS PSTNDPDEFTWTRDSGLVL 60  
QY 85 KTIYVDLFRNGDTSLSLTIEHYISAQAIYVQGISNPSGDLSSGAGLGEPKRWVDETAATGSGV 144  
| : : : : :  
Db 61 KTIYVDLFRNGDTSLSLTIEHYISAQAIYVQGISNPSGDLSSGAGLGEPKRWVDETAATGSGV 120  
QY 145 GRPORDPALRAAMVGFQGMILLDNGYTSATDIYVPLVRNDLSYVAQYWNQTYDLME 204  
| : : : : :  
Db 121 GRPORDPALRAAMVGFQGMILLDNGYTSATDIYVPLVRNDLSYVAQYWNQTYDLME 180  
QY 205 VNGSSFFITIAVQHRALVEGSAFATAVSSCSWCDSQAPILCYLQSFMTGSPILANFDS 264  
| : : : : :  
Db 181 VNGSSFFITIAVQHRALVEGSAFATAVSSCSWCDSQAPILCYLQSFMTGSPILANFDS 240  
QY 265 RSGKDANTLLGSIHTFDEPBAACDDSTFQPCSPRALANHKEVVDSPRSIYTLNDGLSDSEA 324  
| : : : : :  
Db 241 RSGKDANTLLGSIHTFDEPBAACDDSTFQPCSPRALANHKEVVDSPRSIYTLNDGLSDSEA 300  
QY 325 VAVGRYPEDTYNGNPMWFLCTLAABEQLYDALYQMDKQSGLEIVDVSLDFFKALYSDDAAT 384  
| : : : : :  
Db 301 VAVGRYPEDTYNGNPMWFLCTLAABEQLYDALYQMDKQSGLEIVDVSLDFFKALYSDDAAT 360  
QY 385 GTYSSSSSTYSIYDAVKTFADGFVSIYETHAASNGSMSEQYDKSGDEQLSARDLTWSYA 444  
| : : : : :  
Db 361 GTYSSSSSTYSIYDAVKTFADGFVSIYETHAASNGSMSEQYDKSGDEQLSARDLTWSYA 420  
QY 445 ALLTANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSVTWSPSIVA 495  
| : : : : :  
Db 421 ALLTANNRRNSVVPASWGETSASSVPGTCAATSAIGTYSVTWSPSIVA 471

## RESULT 26

ABE80181  
ID ABE80181 standard; protein; 631 AA.

XX AC ABE80181;

XX DT 11-AUG-2003 (first entry)

XX DE A. fumigatus AFGLA3.

XX KM Enzyme; tannase; cellulase; glucose oxidase; glucoamylase; phytase;  
XX KM beta-galactosidase; invertase; lipase; alpha-amylase; laccase;

XX KM polygalacturonase; xylanase; galactate ester linkage; detergent; cellulose;

XX KM glucose; oxygen; myo-inositol phosphate; laccase; tea leaf; sucrose;

XX KM glyceride; starch; maltodextrin; oxidated phenolic compound;

XX KM polygalacturonic acid chain; xylan; xylo-oligomer; food; feed; beverage;  
XX KM textile; tea liquor; cleaning ability.

XX OS Aepergillus fumigatus.

XX PN WO2003012071-A2.

XX PD 13-FEB-2003.

XX PF 05-AUG-2002; 2002WO-US024842.

XX PR 03-AUG-2001; 2001US-0309870P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Jiang B, Storms R, Roemer T, Bussey H;







KW dextrose syrup; syrup; ethanol; fuel; beverage; citric acid;  
 KM ascorbic acid; lysine; glutamic acid.  
 XX Talaromyces emersonii.  
 OS WO9928448-A1.  
 XX  
 XX 10-JUN-1999.  
 PD  
 XX  
 XX 26-NOV-1998; 98MO-DK000520.  
 PF  
 XX  
 XX 26-NOV-1997; 97US-00979673.  
 PR 30-DEC-1997; 97DK-00001557.  
 PR 30-JUN-1998; 98US-00107657.  
 PR 10-JUL-1998; 98DU-00000925.  
 XX  
 XX (NOVO ) NOVO-NORDISK AS.  
 PA  
 XX  
 XX Nielsen BR, Nielsen RI, Lehmbeck J;  
 PI  
 XX  
 XX WPI; 1999-404822/34.  
 DR  
 XX  
 XX New glucamylase obtained from Talaromyces emersonii.  
 PS  
 XX  
 XX Claim 8; Page 57-59; 79pp; English.  
 CC  
 XX  
 XX The present sequence represents a glucamylase enzyme obtained from  
 CC Talaromyces emersonii. The glucamylase enzymes have high thermal  
 CC stability so that a saccharification process may be carried out within a  
 CC shorter period of time or the process may be carried out using a lower  
 CC enzyme dosage. The glucamylase enzymes can be used for saccharifying  
 CC starch hydrolyase for converting starch or partially hydrolysed starch  
 CC into a syrup containing dextrose. They can be used for producing  
 CC oligosaccharides, specialty syrups, ethanol for fuel, beverages or  
 CC organic compounds such as citric acid, ascorbic acid, lysine or glutamic  
 CC acid  
 XX  
 XX  
 SQ Sequence 591 AA;  
 Query Match 60.4%; Score 1674.5; DB 2; Length 591;  
 Best Local Similarity 62.3%; Pred. No. 5.2e-126;  
 Matches 314; Conservative 73; Mismatches 102; Indels 15; Gaps 3;  
 QY 26 TLDWSLNEATVARTALNNGADGAWVSGADGIVVSPSTDPYFYTWTDGGLVK 85  
 DB 4 SLDSPATETPTLQGVNLNIGPNAGADVAGASGIVVSPSRDPTFYTSWTRDALTRAK 63  
 QY 86 TLVDLFRNGDTSILSTIENYISAQAIVOGISNPSGDLSSGAGGEPKFNVDATYTGSG 145  
 DB 64 YLVDAENRGNKDELEQITQYISQAQVQRTISNPSGDLSTG-GIGEPKFNVDATYTGSG 122  
 QY 146 RPRORDPALRATMIGFQGMULDNGTSTRTDIWPELVNRNDLSYVAQYNNQTCYDIWEEY 205  
 DB 123 RPRORDPALRATMIGFQGMULDNGTSTRTDIWPELVNRNDLSYVAQYNNQTCYDIWEEY 182  
 QY 206 NGSSEFTIVQHRALVEGSAFATAVSSGSCWCSOAPELICVQSQWTSFTLANF-DSS 264  
 DB 183 EGSSFTTVAQHRALVEGSAFATAVSSGSCWCSOAPELICVQSQWTSFTLANF-DSS 242  
 QY 265 RSGDKANTLLGSIHTDPPEAACDSTFOPCSPRALNKHKEVDSFRIYLANDGLDSE 324  
 DB 243 RSGKDVNSLIGSIHTDPAGCGDSTFOPCSARALNKHKVYDSFRIYAINSGIAGSA 302  
 QY 325 VAVGRPEDTYKNGNFWFLCTLAARQYDALYQMDKQSLVETDVSLDFKALYSDAAT 384  
 DB 303 VAVGRPEDTYKNGNFWFLCTLAARQYDALYQMDKQSLVETDVSLDFKALYSDAAT 362  
 QY 385 GTVSSSSSTYSIVDVKTFPADGFVSIVETHAASNGSMSEYDKSPGEOLSAADLTW 444  
 DB 363 GTVSSSTYSIVDVKTFPADGFVSIVETHAASNGSMSEYDKSPGEOLSAADLTW 422  
 QY 445 ALLTANRRNSVVPASWGETSASVPGTCATSAIGTSSVTVTSPSIVATGTTTAT 504

DB 423 SLTASARRQSVVPASWGESASVLANVCSATSGATGPTATNTWPS----- 470  
 QY 505 PTGSGVSTSTSKTATYASTSTTT 528  
 DB 471 -SGSGSTTSSAPCTTPPTSVAVT 493  
 RESULT 30  
 AAM51596  
 ID AAM51596 standard; protein; 630 AA.  
 AC AAM51596;  
 XX  
 XX 07-FEB-2002 (first entry)  
 DT  
 XX  
 XX Thielavia terrestris ATCC 20627 glucamylase.  
 DE  
 XX  
 XX Thielavia terrestris; fungus; thermophilic; dextrose; fructose syrup; beer;  
 KW alcohol.  
 XX  
 XX Thielavia terrestris.  
 OS  
 XX  
 XX US6309872-B1.  
 PN  
 XX  
 XX 30-OCT-2001.  
 PD  
 XX  
 XX 01-NOV-2000; 2000US-00704449.  
 PF  
 XX  
 XX 01-NOV-2000; 2000US-00704449.  
 PR  
 XX  
 XX (NOVO ) NOVOZYMES BIOTECH INC.  
 PA  
 XX  
 XX Rey MW, Golightly EJ;  
 PI  
 XX  
 XX WPI; 2002-033282/04.  
 DR N-PSDB; ABA01139.  
 XX  
 XX  
 XX New Thielavia terrestris polypeptide with glucamylase activity, useful  
 PT in the production of dextrose and fructose syrups, beer with low  
 PT carbohydrate content and alcohol from fermentation of raw starch.  
 PS  
 XX  
 XX Claim 5; Fig 1; 42pp; English.  
 CC  
 XX  
 XX The invention relates to an isolated Thielavia terrestris polypeptide  
 CC with glucamylase activity. Thielavia terrestris is a thermophilic  
 CC filamentous fungus that can grow at low pH of 4.5 and elevated  
 CC temperatures of 40-45 degreesC. The glucamylase polypeptide catalyses  
 CC endo/hydrolysis of 1,6-alpha-D-glucoside linkages at points of branching  
 CC chains in 1,4-linked alpha-D-glucose residues. The polypeptide may be  
 CC used in the production of dextrose and fructose syrups, beer with low  
 CC carbohydrate content and alcohol from fermentation of raw starch. The  
 CC present sequence is the polypeptide of the invention  
 CC  
 XX  
 SQ Sequence 630 AA;  
 Query Match 58.3%; Score 1615.5; DB 5; Length 630;  
 Best Local Similarity 58.9%; Pred. No. 3.3e-121;  
 Matches 315; Conservative 83; Mismatches 112; Indels 25; Gaps 7;  
 QY 6 LIALSLVCTGLANVISKRA-----TLDWSLNEATVARTALNNGADGAWVSGA 56  
 DB 6 LIGLALLLPAALGHPASRVREGEVRSVDSFATESPILSNLNCIGSGCHASGV 65  
 QY 57 DSGIVVSPSTNDPDTFYTWTDGGLVLTVDLPNG-DTSLSTIENYISAQAIVOGI 115  
 DB 66 ASGIVVSPSTNDPDTFYTWTDGGLVLTVDLPNG-DTSLSTIENYISAQAIVOGI 125  
 QY 116 SNPSGDLSSGAGGEPKFNVDATYTGSGRPRORDPALRATMIGFQGMULDNGTSTA 175  
 DB 126 SNPSGDLSSGAGGEPKFNVDATYTGSGRPRORDPALRATMIGFQGMULDNGTSTA 185  
 QY 176 TDIWPELVNRNDLSYVAQYNNQTCYDIWEEYSSFTTIVQHRALVEGSAFATAVSSG 235



Db 186 SSIWPIYKNDLAVVAQ--NNTGFDLMBEYSGSSFFVNAOHRAVLVEGALLATSLGTS 243  
 QY 236 WCDSDQAEILCYLQSFMTGS--FLIANFDSSRGKDNATLLGSIHTPEACDDSTFOP 293  
 Db 244 ACSAVAPQILCFLOSFSPSSGYLIAN----STAKDANTLLGSIHTPEACDDAATFOP 299  
 QY 294 CSRRALANHEVVDSPFSITTLNDGLSDSEAVAVGRPEPTTYNGNWFELCTAAAEOLY 353  
 Db 300 CSBRALANHEVVDSPFSITTLNDGLSDSEAVAVGRPEPTTYNGNWFELCTAAAEOLY 359  
 QY 354 DALYOMDKGSLLEVTDVSLDFPKALYSDAATGTYSSSSSTYSIVDAVKTFADQFVSIVE 413  
 Db 360 DALYOMDKGSLLEVTDVSLDFPKALYSDAATGTYSSSSSTYSIVDAVKTFADQFVSIVE 419  
 QY 414 THAASNGSMSEQYDKSDGQSLASADLTWSYAAALLTANNRNSVVPASWGETSASVPGTC 473  
 Db 420 QYAGTQNTSLSEQFSKTGEPFLSAVDLTWSYAAALLTANNRNSVVPASWGETSASVPGTC 479  
 QY 474 AATSAICTGSSVTVTSPSIVATGCTTTATPTGSGSVTSTKTTATASKTSTTT 528  
 Db 480 SATSVGSSYTSATATSPFP-----SQTPTSSTGAGS-SPASSTTATATACSTPT 527

RESULT 31

AA71034  
 ID AAR71034 standard; protein; 626 AA.

XX AAR71034;

DT 25-MAR-2003 (revised)  
 DT 18-MAR-1996 (first entry)

XX N. crassa glucanase.

KM Neurospora crassa; glucanase gene; gla-1; promoter;  
 KM expression construct; primer.

XX Neurospora crassa.

XX Key Location/Qualifiers

FT Peptide 1..19

FT Protein /note= "signal peptide" 20..626

FT Misc-difference 27..28 /note= "mature gla-1"

FT Misc-difference 34..35 /note= "propeptide processing site"

XX /note= "propeptide processing site"

XX W09505474-A2.

XX 23-FEB-1995.

XX 15-AUG-1994; 94WO-GB001789.

XX 13-AUG-1993; 93GB-00016883.

XX (UYLE-) UNIV LEBDS.

XX Radford A, Parish JH;

XX WPI, 1995-098771/13.

XX N-PSDB; AA084689, AA084695.

XX Regulated glucanase promoter - useful for producing heterologous

XX polypeptide(s) in filamentous fungi.

XX Claim 1, Fig 1; 36pp; English.

XX This sequence represents N. crassa glucanase. The gene encoding this

XX sequence has high expression and is one of the major secreted proteins of

XX N. crassa when starch-induced. The ORF of this gene and the promoter

XX region can be attached in frame to a foreign gene thereby introducing

CC foreign proteins into the N. crassa secretory pathway. The 5' primer  
 CC (AA084690) encompassed the unique PvuII site at posn. 2163 of the  
 CC encoding sequence and the 3' primer (AA084691) conty. an MroI site  
 CC hybridises at the 3' end of the gla gene. The 5' upstream PCR fragment  
 CC was amplified and cloned into the SmaI site in a pNB 193 vector. The  
 CC clone was named pMro. The remainder of the gla gene was inserted by  
 CC digestion of the glucanase clone pGla-XhoI (AA084692). This plasmid  
 CC contains the entire gla gene however the downstream unsequenced and non-  
 CC transcribed area was deleted. pGla-XhoI was digested with SacI and PvuII  
 CC and the fragment ligated into the SacI/PvuII sites of pMro. The SacI  
 CC site of pGla-XhoI was derived from the linker and not from the coding  
 CC region of glucanase consequently, no glucanase sequence was deleted  
 CC (see AA084693, pGla-MroI). In an attempt to increase transcription  
 CC efficiency, 1575 bp were deleted from the encoding sequence, creating the  
 CC plasmid pG6 (plasmid glucanase, EcoRI). Deposits of plasmids pGla-Xho  
 CC (AA084692), pGla-MroI (AA084693), pG6 (AA084694) have been made and the  
 CC deposition details are to be added to the patent application. (Updated on  
 CC 25-MAR-2003 to correct PN field.)  
 XX  
 XX Sequence 626 AA;

Query Match 56.9%; Score 1577; DB 2; Length 626;

Best Local Similarity 57.6%; Pred. No. 4,1e-118;

Matches 306; Conservative 85; Mismatches 130; Indels 10; Gaps 6;

QY 2 SFRSLALSGLVCTGLANVISKRAITDLSWLSNEATVARTAILNNITAGDAMWGSAGSY 61

Db 13 AFQAVLGLPDLPEKHSDIRK-SVDSYIOTETPIAOKNLCNIGASGRASGAGV 71

QY 62 VASPSNDPDYPTVTRDGLVLYKTVLDFRNG-DTSLSTIENYISAQIVGGINPSG 120

Db 72 VASPSKSPDYMTYTRDALVTKLVDFEFTNDYNTLIDVTDIAVAQKLGAVNPSG 131

QY 121 DLSGAGLGEPKENVDETAAYTSGWGRPORDGALRAATAMIGFQWLLDNGYSTADIW 180

Db 132 SLNNGAGLEPKRMVLDQFTGAMGRPQDGPRLRALIGYKWLVSNGYADTAISIT 191

QY 181 PLVRNDLSVYAQYNTGTDLMEVNGSSFTTAVOHRALVEGAPATAVGSSCWDCSO 240

Db 192 PIVKNDLAVTAQYWNNTGFDLMEVNSSFFTTAASHRALVEGASAPAKSGSSCSACDAI 251

QY 241 APRILCYLQSFMTGS--FLIANFDSSRGKDNATLLGSIHTPEACDDSTFOPCSBRAL 299

Db 252 APQILCFLOSFMSNGSITISNFVNRSGDINSVLISHNFDPAACGDVNTFPCCSDRAL 311

QY 300 ANHEVVDSPFSIYTLNDGLSDSEAVAVGRPEPTTYNGNWFELCTAAAEOLYDALYON 359

Db 312 ANHKVVDSPFSIYTLNDGLSDSEAVAVGRPEPTTYNGNWFELCTAAAEOLYDALYON 370

QY 360 DRQSLLEVTDVSLDFPKALYSDAATGTYSSSSSTYSIVDAVKTFADQFVSIVETHAASN 419

Db 371 KRQSLLEVTDVSLDFPKALYSDAATGTYSSSSSTYSIVDAVKTFADQFVSIVETHAASN 430

QY 420 GSWSEQYDKSDGQSLASADLTWSYAAALLTANNRNSVVPASWGETSASVPGCSASAI 479

Db 431 GSIAEQFDKSDGAPLSATHLTWSYASFLSAAARAGIIVPSWGAANSIPGCSASTVA 490

QY 480 GYSSVTVTSPSIVATGCTTT-ATPTGSGS-----VTSTSKTTATASKT 524

Db 491 GSATATATSPSIVATGCTTT-ATPTGSGS-----VTSTSKTTATASKT 541

RESULT 32

AA71034

XX AAB18823

XX AAB18823;

XX 08-FEB-2001 (first entry)

XX Amino acid sequence of a glucanase polypeptide.

XX Glucanase; protein production; promoter; hormone; receptor; antibody.

XX	OS	Fusarium venenatum.
XX	FH	Key
XX	FT	Location/Qualifiers
XX	FT	1..21
XX	FT	/note= "signal peptide"
XX	PN	MO200056900-A2.
XX	XX	
XX	PD	28-SEP-2000.
XX	PF	22-MAR-2000; 2000MO-US007815.
XX	PR	22-MAR-1999; 99US-00274449.
XX	PA	(NOVO ) NOVO NORDISK BIOTECH INC.
XX	PI	Berka RM, Rey MW, Brown K, Brown SH;
XX	DR	WPI; 2000-638265/61.
XX	DR	N-PSDB; AAA75945.
PT	PT	Promoters useful for expressing heterologous genes and producing
PT	PT	polypeptides such as hormones, receptors, antibodies or enzymes in a
PT	PT	fungus cell.
PS	PS	Example 8; Fig 1A-F; 104pp; English.
XX	XX	
CC	CC	The present sequence represents a glucanase polypeptide. The promoter
CC	CC	sequence of this gene is useful for producing a polypeptide, preferably a
CC	CC	hormone, receptor, antibody, a reporter or an enzyme selected from
CC	CC	oxidoreductase, transferase, hydrolase, lyase, isomerase or ligase, in
CC	CC	particular amnopolypeptide, amylase, carbohydrase, carboxypeptidase,
CC	CC	calase, cellulase, chitinase, cutinase, cyclodextrin glycosyl
CC	CC	transferase, deoxyribonuclease, esterase, alpha-galactosidase, beta-
CC	CC	galactosidase, glucanase, alpha-glucosidase, beta-glucosidase,
CC	CC	invertase, lactase, lipase, mannosidase, mutase, oxidase, a
CC	CC	pectinolytic enzyme, peroxidase, phytase, polyphenoloxidase, proteolytic
CC	CC	enzyme, ribonuclease, transglutaminase or xylanase, in fungal host cells
XX	XX	
SQ	SQ	Sequence 581 AA;
		Query Match 51.4%; Score 1425.5; DB 3; Length 581;
		Best Local Similarity 52.2%; Pred. No. 6.2e-106;
		Matches 273; Conservative 81; Mismatches 146; Indels 23; Gaps 6;
QY	9	LSGLVCTGL--ANVT---SKRATLDSWLSNEATVARTLAINNIGADGAMVSGADSGIYVA 63
DB	6	LYGLVAAALMQGVAAAPSKDNLSEFRIDKADISIKGLVANIADGKRAQAGAPAAVVA 65
QY	64	SPSTNDPNPYFYWTWRDSCGLVLTVDLFRANGDPTSLSTIENTYISAQAVIGISNFGDLS 123
DB	66	SPSKNDPPYWTWTRDSALTYKLVVERFIHQDKSLQRKIDEVSAQAKLQGTTFSGSPE 125
QY	124	SGAGLGEPEKENVDEFAVYGSWGRPPRDGPALPALTAMIGFGQWLNDNGYTTATTDIYVPLV 183
DB	126	SG-GLGEPEKFHYNLTAFTGSGWGRPPRDGPPLATATLTLYAEWLISHGERSKALNTWPEVI 184
QY	184	RNDLSVYAQYNNQGYDILMEEVNGSSFFLIANQHVALVGSFAFNAVSSGSCWCOSQAP 243
DB	185	EKDLATYTKKFMKRTYDILMEEVNGSSFFLSHSHALVVGALAAKLKDGKSCDDCVTNMR 244
QY	244	ILCYLQSFMTGSFIILANFP--SSRSKGDANTLGSIHTEFPEMACDDSTFGCSPRALAN 301
DB	245	VLFCFQITFWTGSVYDNNINVKGRGLDYNSTLSSIHFPDPSKCTDSTFGCSPRALAN 304
QY	302	HKEVVDTSRSIYTLNDGLSDSEAVAIVGRYPEDTYNNGNFWELCTAAAEQLYDALYQMDK 361
DB	305	HKAUVDSFHSISGVAKNMGQKAAAVGRYSEDEVYDNGNFWELATLAAAEQLYAAVYQMDK 364
QY	362	QGLSEVTVYSLDFEFKALVSDATGTYSSSSSTYSSSIIVA VKTFADGPFYSIVETHAASNGS 421
DB	365	LGAVTVDDVSLSFEPKQIVPKVSKTYAKTKYKILKAKTYADGPFVAAVVDTYTPKQGS 424

[illegible]

```

Db      6 LVGLVASALMQGVVAVSPSKDNLERFIDKQADISIKGVLANIGADGKRAQGAAPAVVA 65
Qy      64 SPSTNDPDEYFTYTRSGVLVLTVDLFRNGDTSLSSTIENYISAQAIVOGISNPSGDL 123
        66 SPKEDPDYWTYTRBSALTYYKVLVERFHGDKSLQKRIKEYSAQAKIGTTNPPSGSPE 125
Qy      124 SGAGLGEPRKNNVDETATYTGSGWRPQDRGPRALRATAMIGFQWMLDNGYSTATDIWPLV 183
        126 SG-GLGEPRKPHVNLTFGTGSMGRPQDRGPRALRATALTVAEWLISHGERSKALNKWPVI 184
Db      184 RNLSVYAQVWNTGYDLMEENVNGSSFFTLAVOHRALVESAFAPTAVNGSSCSDQAP 243
        185 EKDLATYTKRWNRKGYDLMEENVNGSSFFTLASHRALVBSGALAKKLGKSCPCVTNA 244
Qy      244 ILCTYLOSFWTGSFTLANFD--SSRSGKDANTLGSIHTEPDPEACDDSTFQPCSPRALAN 301
        245 VLCTFLOTFTGTGYYDVSNINVKDGRKGLDVNSILSSIHTEPDNSKCTDSTFQPCSPRALAN 304
Qy      302 HKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPFVLTCTLAABQLYDALYQMDK 361
        305 HKAIVVDSFRSIYGVNKNRGQKAAVGRYSEDYVDGNPMVYATLAAAEOLYAAVYQMDK 364
Db      362 QGSLLEVTVSLDPEFKALYSDAATGTYSSSSSTYSIVDAVKTADGFVSIIVETHAASNGS 421
        365 LGAVTVDDVSLSPFKDIVPVKSGTYAKTKYKEIKAKTYADGFVAAVVQYTPRKDGS 424
Qy      422 MSBOYDKSDGEOLASDRLTWSYALLTANNRRNSVVPASWGETSASSVPGETCAATSAIGT 481
        425 LABQFDKSTGAPKSAVHLTWSYAAFVATTERBDGIIISPSWGESANKNVPAVCAAPACDT 484
Qy      482 -----YSSVTVTWSPSIVATGTTT-----TPTGSG 509
        485 TITFSVKNVQVSSDQKVVVGSVTLELNSMSPDDGIALTPSSSG 527
Db

```

RESULT 34

ABO01925

ID ABO01925 standard; protein; 581 AA.

AC ABO01925;

XX

DT 08-AUG-2003 (first entry)

XX

DE Fusarium venenatum glucoamylase.

XX

KW Promoter; glucoamylase; Daria; Quim; heterologous gene expression.

XX

OS Fusarium venenatum.

XX

PN US6518044-B1.

XX

PD 11-FEB-2003.

XX

PF 30-OCT-2001; 2001US-00999201.

XX

PR 22-MAR-1999; 99US-00274449.

XX

PR 22-JUL-1999; 99US-0145339P.

XX

PR 22-MAR-2000; 2000US-00534407.

XX

PA (NOVO) NOVOZYMES BIOTECH INC.

XX

PI Berka RM, Rey MM, Brown K, Brown SH;

XX

DR N-PSDB; ACD07856.

XX

XX

PT Producing a polypeptide through the use of promoters that express genes

PT in fungal cell, comprises cultivating a fungal host cell in a medium for

PT the production of the polypeptide and isolating the polypeptide from the

PT cultivation medium.

XX

XX

PS Example 7, Fig 2; 57pp; English.

XX

```

CC      CC The invention relates to producing a polypeptide comprises cultivating a
CC      CC fungal host cell in a medium for the production of the polypeptide and
CC      CC isolating the polypeptide from the cultivation medium. The host cell
CC      CC comprises a first nucleic acid sequence encoding the polypeptide operably
CC      CC linked to a second nucleic acid sequence having a promoter foreign to the
CC      CC first sequence. The promoter comprises a nucleic acid sequence: (a)
CC      CC having nucleotides 1-938 of the Fusarium venenatum Daria gene appearing
CC      CC as ACD07859; or (b) that retains the promoter activity of nucleotides 1-
CC      CC 938 of the Daria gene; or (c) that hybridises under medium stringency
CC      CC conditions with nucleotides 1-938 of the Daria gene. Also disclosed are
CC      CC the F. venenatum Daria protein (a novel secreted protein), the F.
CC      CC venenatum Quim gene/protein (a vacuolar associated protein) and the F.
CC      CC venenatum glucoamylase gene/protein. The method is useful for producing a
CC      CC polypeptide in commercially relevant quantities by using promoters that
CC      CC express genes in fungal cells. The present sequence represents one of the
CC      CC disclosed F. venenatum proteins
CC      CC
SQ      Sequence 581 AA;

```

Query Match 51.4%; Score 1425.5; DB 7; Length 581;

Best Local Similarity 52.2%; Pred. No. 6, 2e-106;

Matches 273; Conservative 81; Mismatches 146; Indels 23; Gaps 6;

```

Qy      9 LSGIVCTGL--ANVT---SKRATLDSWLSNEARTVARTALINIGACAWGSGADSGIVVA 63
        6 LVGLVASALMQGVVAVSPSKDNLERFIDKQADISIKGVLANIGADGKRAQGAAPAVVA 65
Db
Qy      64 SPSTNDPDEYFTYTRSGVLVLTVDLFRNGDTSLSSTIENYISAQAIVOGISNPSGDL 123
        66 SPKEDPDYWTYTRBSALTYYKVLVERFHGDKSLQKRIKEYSAQAKIGTTNPPSGSPE 125
Db
Qy      124 SGAGLGEPRKNNVDETATYTGSGWRPQDRGPRALRATAMIGFQWMLDNGYSTATDIWPLV 183
        126 SG-GLGEPRKPHVNLTFGTGSMGRPQDRGPRALRATALTVAEWLISHGERSKALNKWPVI 184
Db
Qy      184 RNLSVYAQVWNTGYDLMEENVNGSSFFTLAVOHRALVESAFAPTAVNGSSCSDQAP 243
        185 EKDLATYTKRWNRKGYDLMEENVNGSSFFTLASHRALVBSGALAKKLGKSCPCVTNA 244
Db
Qy      244 ILCTYLOSFWTGSFTLANFD--SSRSGKDANTLGSIHTEPDPEACDDSTFQPCSPRALAN 301
        245 VLCTFLOTFTGTGYYDVSNINVKDGRKGLDVNSILSSIHTEPDNSKCTDSTFQPCSPRALAN 304
Db
Qy      302 HKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPFVLTCTLAABQLYDALYQMDK 361
        305 HKAIVVDSFRSIYGVNKNRGQKAAVGRYSEDYVDGNPMVYATLAAAEOLYAAVYQMDK 364
Qy      362 QGSLLEVTVSLDPEFKALYSDAATGTYSSSSSTYSIVDAVKTADGFVSIIVETHAASNGS 421
        365 LGAVTVDDVSLSPFKDIVPVKSGTYAKTKYKEIKAKTYADGFVAAVVQYTPRKDGS 424
Db
Qy      422 MSBOYDKSDGEOLASDRLTWSYALLTANNRRNSVVPASWGETSASSVPGETCAATSAIGT 481
        425 LABQFDKSTGAPKSAVHLTWSYAAFVATTERBDGIIISPSWGESANKNVPAVCAAPACDT 484
Qy      482 -----YSSVTVTWSPSIVATGTTT-----TPTGSG 509
        485 TITFSVKNVQVSSDQKVVVGSVTLELNSMSPDDGIALTPSSSG 527
Db

```

RESULT 35

ABB80170

ID ABB80170 standard; protein; 704 AA.

AC ABB80170;

XX

DT 11-AUG-2003 (first entry)

XX

XX

DE A. fumigatus AFG1.

XX

KW Enzyme; tannase; cellulase; glucose oxidase; glucoamylase; phytase;

KW beta-galactosidase; invertase; lipase; alpha-amylase; laccase;

KW polygalacturonase; xyranase; galactate ester linkage; detergent; cellulose;

KW glucose; oxygen; myo-inositol phosphate; lactose; tea leaf; sucrose;  
 KW glyceride; starch; maltodextrin; oxidized phenolic compound;  
 KW polygalacturonic acid chain; xylan; xylo-oligomer; food; feed; beverage;  
 KW textile; tea liquor; cleaning ability.  
 XX Aspergillus fumigatus.  
 OS WO2003012071-A2.  
 PN 13-FEB-2003.  
 XX 05-AUG-2002; 2002WO-US024842.  
 XX 03-AUG-2001; 2001US-0309870P.  
 XX (ELIT-) ELITRA PHARM INC.  
 PA Jiang B, Storms R, Roemer T, Bussey H;  
 PI WPI; 2003-332729/31.  
 DR N-PSDB; ABO80331, ABO80332.  
 XX Novel isolated Aspergillus fumigatus polypeptide, useful in various  
 PT industries such as those involved in the making of food and feed,  
 PT beverages, textiles and detergents.  
 PT Claim 17; Page 100-01, 169pp; English.  
 XX The sequences given in ABB80164-87 show enzymatic proteins derived from  
 CC A. fumigatus. These proteins display the catalytic activity of an enzyme  
 CC such as tannase, cellulase, lipase, alpha-amylase, laccase, beta-  
 CC galactosidase, invertase, lipase, alpha-amylase, laccase,  
 CC polygalacturonase or xylanase. Compositions comprising the tannase are  
 CC useful for modulating the amount of compounds that comprise a galacte  
 CC ester linkage in a composition. Compositions comprising cellulase are  
 CC useful for modulating the amount of cellulose in a composition.  
 CC Compositions comprising glucose oxidase are useful for modulating the  
 CC amount of glucose or oxygen in a composition. Compositions comprising  
 CC phytase are useful for modulating the amount of myo-inositol phosphates  
 CC in a composition. Compositions comprising beta-galactosidase are useful  
 CC for modulating the amount of lactose in a composition. Compositions  
 CC comprising sucrose or invertase are useful for modulating the amount of  
 CC sucrose in a composition. Compositions comprising lipase are useful for  
 CC modulating the amount of glyceride in a composition. Compositions  
 CC comprising alpha-amylase are useful for modulating the amount of  
 CC starches or maltodextrins in a composition. Compositions comprising  
 CC laccase are useful for modulating the amount of oxidized phenolic  
 CC compounds in a composition. Compositions comprising polygalacturonases  
 CC are useful for modulating the amount of high or low molecular weight  
 CC polygalacturonic acid chains in a composition. Compositions comprising  
 CC xylanases are useful for modulating the amount of xylan or xylo-oligomers  
 CC in a composition. The A. fumigatus proteins and corresponding DNA's are  
 CC useful in various industries such as those involved in the making of food  
 CC and feed, beverages, textiles and detergents. The DNA's are useful to  
 CC express recombinant enzymes for characterization, modification or  
 CC industrial uses, to compare with the nucleotide sequence of A. fumigatus  
 CC to identify duplicated genes of paralogs having the same or similar  
 CC biochemical activity and/or function, to compare with nucleic acid  
 CC sequence of other related or distant fungal organisms to identify  
 CC potential orthologous enzyme genes, for selecting and making oligomers  
 CC for attachment to a nucleic acid array for examination of expression  
 CC patterns, and to raise anti-protein antibodies. The polypeptide having  
 CC tannase activity increases the yield of tea liquor from tea leaves,  
 CC improves the colour, flavour and health benefits of tea products,  
 CC particularly an instant tea product. The polypeptide having cellulase  
 CC activity enhances cleaning ability of detergent compositions  
 CC  
 XX Sequence 704 AA:

Query Match 50.8%; Score 1408.5; DB 6; Length 704;  
 Best Local Similarity 52.0%; Pred. No. 1.9e-104;  
 Matches 268; Conservative 85; Mismatches 139; Indels 23; Gaps 5;

QY 27 LDWLSNEATVARTAIINNIGADGAWSGADSGIVVASTNDPFIYTWTDGSLVLT 86  
 DB 118 LWSLQAQETSVALDGVNINVGPNKAKATGASGIIITIASPSQNPDDVYTWTDALTVKY 177  
 QY 87 LYDLF-RNGDTSLLSTIENYISAQAVGCSINPSGSLSGALGEEKFNVDEATAYGSMG 145  
 DB 178 LVGSFAADHPAIQRIIEDYVESQAHLOTVSNPSGLSSG-GLGEKFLVVDGSAFPGSWG 236  
 QY 146 RPRDGPALRATAMIGFGQMLDNGYSTATDIIWPLVNDLSYVAQYNNQGTGDIAMEEV 205  
 DB 237 RPOSQDPALRATTLISYALMIDNGYFSTVESIWIPIQNDLSYLFERNSSFTDIMEEV 296  
 QY 206 NGSSEFTTAVGHRALVESAFAFVAVGSSCWCDGQAPILCYLOSFWTGSPFIANFDSR 265  
 DB 297 RGSSEFTTAVGHRALVESAFAFVAVGSSCWCDGQAPILCYLOSFWTGSPFIANFDSR 356  
 QY 266 SGKDANTLLGSIHTPPEAACDSTFQPCSPALANHKRVDSFRSITYLNDGSLSEAV 325  
 DB 357 SGKDANSILGIHTPDPNAGCDGQTFQPCSDRALSHKEVDSFRSLYPMADIPOGQAV 416  
 QY 326 AVGRYPEDTYNGNPMFLCTLAAAEQYDALYQMDKQSLSEYTVDELDFPKALYSPAATG 385  
 DB 417 AVGRYPEDTYNGNPMFLCTLAAAEQYDALYQMDKQSLSEYTVDELDFPKALYSPAATG 476  
 QY 386 TYSSSSSTYSIVDAVKTFADGFVSIVETHAASNGSMSBOYKSDGEOUSARDLWYSYA 445  
 DB 477 TYAKDITFASISAAVRDYADRFLRVQKTPPNGLAQFGRYDGSPLSAQDLTYSYS 516  
 QY 446 LITA-----NRRNSVVPASWGETSASVPGTAAISAGTYSVTVSWP----- 491  
 DB 537 FLTAQVARRHALNPSASHIOPILSNATYTPALPOVCTPSSARGPYOPVRKIKRPRECS 596  
 QY 492 --SIVATGGTTTAAPTGS-----GSVTSKTTA 519  
 DB 597 PRSTVAVRNVLTATYIGEDIFLVGSI PALGEMDA 631

RESULT 36

ABB80174  
 ID ABB80174 standard; protein; 620 AA.

XX ABB80174;

DT 11-AUG-2003 (first entry)

XX A. fumigatus AFGLA2.

XX Enzyme; tannase; cellulase; glucose oxidase; glucoamylase; phytase;  
 KW beta-galactosidase; invertase; lipase; alpha-amylase; laccase;  
 KW polygalacturonase; xylanase; gallate ester linkage; detergent; cellulose;  
 KW glucose; oxygen; myo-inositol phosphate; lactose; tea leaf; sucrose;  
 KW glyceride; starch; maltodextrin; oxidized phenolic compound;  
 KW polygalacturonic acid chain; xylan; xylo-oligomer; food; feed; beverage;  
 KW textile; tea liquor; cleaning ability.

XX Aspergillus fumigatus.

OS WO2003012071-A2.

PN 13-FEB-2003.

XX 05-AUG-2002; 2002WO-US024842.

XX 03-AUG-2001; 2001US-0309870P.

XX (ELIT-) ELITRA PHARM INC.

PA Jiang B, Storms R, Roemer T, Bussey H;

PI WPI; 2003-332729/31.

DR N-PSDB; ABO80339, ABO80340.

PT Novel isolated Aspergillus fumigatus polypeptide, useful in various

PT industries such as those involved in the making of food and feed,  
PT beverages, textiles and detergents.

PS Claim 17, Page 123-24; 169pp; English.

XX  
CC The sequences given in ABB80164-87 show enzymatic proteins derived from  
CC A. fumigatus. These proteins display the catalytic activity of an enzyme  
CC such as tannase, cellulase, glucose oxidase, glucoamylase, phytase, beta-  
CC galactosidases, invertase, lipase, alpha-amylase, lactase,  
CC polygalacturonase or xylanase. Compositions comprising the tannase are  
CC useful for modulating the amount of compounds that comprise a gallate  
CC ester linkage in a composition. Compositions comprising cellulase are  
CC useful for modulating the amount of cellulose in a composition.  
CC Compositions comprising glucose oxidase are useful for modulating the  
CC amount of glucose or oxygen in a composition. Compositions comprising  
CC phytase are useful for modulating the amount of myo-inositol phosphates  
CC in a composition. Compositions comprising beta-galactosidases are useful  
CC for modulating the amount of lactose in a composition. Compositions  
CC comprising sucrose or invertase are useful for modulating the amount of  
CC sucrose in a composition. Compositions comprising lipase are useful for  
CC modulating the amount of glyceride in a composition. Compositions  
CC comprising alpha-amylases are useful for modulating the amount of  
CC starches or malto-dextrins in a composition. Compositions comprising  
CC lactase are useful for modulating the amount of oxidized phenolic  
CC compounds in a composition. Compositions comprising polygalacturonases  
CC are useful for modulating the amount of high or low molecular weight  
CC polygalacturonic acid chains in a composition. Compositions comprising  
CC xylanases are useful for modulating the amount of xylan or xyl-o-oligomers  
CC in a composition. The A. fumigatus proteins and corresponding DNA's are  
CC useful in various industries such as those involved in the making of food  
CC and feed, beverages, textiles and detergents. The DNA's are useful to  
CC express recombinant enzymes for characterization, modification or  
CC industrial uses, to compare with the nucleotide sequence of A. fumigatus  
CC to identify duplicated genes of paralogs having the same or similar  
CC biochemical activity and/or function, to compare with nucleic acid  
CC sequence of other related or distant fungal organisms to identify  
CC potential orthologous enzyme genes, for selecting and making oligomers  
CC for attachment to a nucleic acid array for examination of expression  
CC patterns, and to raise anti-protein antibodies. The polypeptide having  
CC tannase activity increases the yield of tea liquor from tea leaves,  
CC improves the colour, flavour and health benefits of tea products,  
CC particularly an instant tea product. The polypeptide having cellulase  
CC activity enhances cleaning ability of detergent compositions

CC  
XX  
SQ Sequence 620 AA;

Query Match 49.8%; Score 1380.5; DB 6; Length 620;

Best Local Similarity 53.8%; Pred. No. 2.9e-102; Mismatches 267; Conservative 75; Indels 23; Gaps 3;

QY 13 VCTGLANVISKRAAT-----LDSTWLSNEATVATATLNNIGADGAWVSGADS 58  
DB 6 IAVGIASVLGHLATPSSAKELAAASPAANSRLVKEGFANESILALNGTINAPPTA 65  
QY 59 GIVVASPESTNDPFFYTWTRDGLVLTVDLPFNQGT-----SLISTENTYSAQAI 111  
DB 66 GLTIASPTIQNPDFYTWTRDAALTFKGLVDIFGCGTFFIVNLDGETHIQDIYISSQAV 125  
QY 112 VQGISNSGDISAGAGEPEKENVDETAYGSMGRPORDPALATMIGCGQLNGY 171  
DB 126 LQNSNPSGRLSDSGAGEPEKENVDETAYGSMGRPORDPALATMIGCGQLNGY 185  
QY 172 TSTATDVLWPLVRNDLSYVAQYNNQGTDLWEENVGSSFFPIAVQHPALVEGSAFAAVG 231  
DB 186 QSVASNLIMPVANDLITVAQYNNHTGDFDLWEELIDGSSFFITVAQHAMEGSAIAQALG 245  
QY 232 SSCGWCDSQADELICYIQSFWTGSPFIANF--DSRSGKDNLTGSIHTFPDPAACDS 289  
DB 246 KPHAGYDAVAPEILICLLQSYNNESALISININNNRGSGIDINSVLTSTHTFDPAACDS 305  
QY 290 TFGQCSRRALANHEVVDFFRSITLNDGLSDSEAVAVGRPEPTYYNNGNWFCTLAAA 349  
DB 306 TFGQCSRRALANHEVVDFFRSITLNDGLSDSEAVAVGRPEPTYYNNGNWFCTLAAA 365

QY 350 EQLYDALYQWQKQGSLEVTDSIDFFKALYSDAATGTSSSSSTYSIYDAVKTPADGV 409  
DB 366 EQLYDALYQWQKQGYLTVTQTSIAFFRPSSTVEPTGYKNTFNPKSLTEYRTVADDF 425  
QY 410 SIYETHAANGSMSEQYDSDGEQASARLTWSYALTLANNRRNSVAPSWGETASAV 469  
DB 426 FLVEKTPSPNGSLAEYDRNTGVLSANDLTWSYAAFLSTLORLIMPDSWGPSSANTV 485  
QY 470 PGTCAATSAIGTVSSV 485  
DB 486 PTTCSKTITITGYSAV 501

RESULT 37  
AAWJ0155  
ID AAWJ0155 standard; protein; 616 AA.

XX AAWJ0155;  
XX 17-OCT-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 15-JAN-1998 (first entry)  
XX Glucoamylase P.  
XX glucoamylase P; hormoconis resinae; debranching activity; enzyme;  
KM dextrinase activity; Trichoderma reesei; starch granule; preservation;  
KM hydrolysis; gelatinised starch; glucose syrup; straight linear dextrin;  
KM saccharification; lignocellulosic material; sugar utilisation;  
KM feed additive.

XX Amorphotheca resinae.

PN US565585-A.

PD 09-SEP-1997.

PF 07-FEB-1995; 95US-00385370.

XX 03-SEP-1992; 92US-00937789.

PR 12-AUG-1993; 93US-00104853.

XX (ALKO-) ALKO-YHTIOT OY.

XX Jouteloki V, Vainio A, Fagerstrom R, Nevalainen H, Aho S;

PI Torkkeli T, Korhola M, Torkkeli H;

XX WPI; 1997-456802/42.

DR N-PSDB; AAT90830, AAT90831.

XX Hormoconis resinae glucoamylase P gene construct - for transforming  
PT Trichoderma to produce recombinant glucoamylase P.

PS Claim 1; Col 58-60; 61pp; English.

XX This sequence represents the Hormoconis resinae glucoamylase P. The DNA  
CC encoding this sequence is used in the composition of the invention and  
CC are capable of being processed by a Trichoderma host cell. H. resinae  
CC glucoamylase P has higher debranching and dextrinase activity than  
CC conventional glucoamylase-pullulanase mixtures. T. reesei secretes  
CC enzymes that are important for the degradation of complexes around and in  
CC starch granules. The recombinant glucoamylase P enzyme produced by the  
CC Trichoderma host cell is useful in applications requiring the hydrolysis  
CC of gelatinised starch or the presence of a debranching activity. These  
CC applications include starch analysis, the manufacture of glucose syrups,  
CC production of straight linear dextrin for use in food, medicines and  
CC cosmetics, and in the preparation of food fibers by the enzymatic  
CC treatment of seed husks or brans. The enzyme can also be used as an  
CC additive to laundry and dish washing detergents, in wood and textile  
CC industry applications such as the preparation of lignocellulosic materials.  
CC It can also be used in the saccharification of animal or vegetable fodder, the  
CC preservation of protein-containing animal or vegetable fodder, the



```

AAP60359
ID AAP60359 standard; protein; 604 AA.
AC AAP60359;
XX
XX
DT 23-MAY-1991 (first entry)
DE Glucoamylase structural gene product.
XX
XX Protease; Rhizopus; amylase; alcohol.
XX
XX Rhizopus oryzae.
OS
FH Key Location/Qualifiers
FT Peptide 1..25
FT /label= signal peptide
FT Protein 26..604
FT /label= Mature peptide
XX
XX EPI86066-A.
PN 02-JUL-1986.
XX
XX 13-DEC-1985; 85EP-00115910.
XX
XX 15-DEC-1984; 84UP-00264964.
XX
XX (SUNR ) SUNTORY LTD.
XX
XX Aahihari T, Nakamura N, Tanaka Y, Shibano Y, Yoshizumi H;
XX WPI, 1986-170627/27.
XX N-PSDB; AAN60292.
XX
XX Gluco-amylase gene obd. from rhizopus strain - useful in transforming
XX yeasts or bacteria for efficient prodn. of glucoamylase for starch
XX hydrolysis in alcohol prodn.
XX
XX Claim 3; Fig 1; 108pp; English.
XX
XX The gene may be used to form vectors capable of transforming an E.coli or
XX B.subtilis expression sysem for the economical production of the enzyme.
XX Glucoamylase has a good enzymatic activity for hydrolysing raw starch
XX especially in the manufacture of alcohol, and is not degraded by protease
XX activity during production
XX
XX Sequence 604 AA;
SQ
Query Match 26.1%; Score 724; DB 1; Length 604;
Best Local Similarity 36.8%; Pred. No. 2.6e-49;
Matches 165; Conservative 71; Mismatches 170; Indels 42; Gaps 10;
QY 25 ATLDSWLNSNATYARPTAILNIGADGAWSGADSGIVVASTPDNDPFFYTWTRDGLVL 84
DB 168 STISSWTKKQEGISRFMLKRNINP-----PGSATGFPAASLSTGPDYTYAMTRDALITS 222
QY 85 KTLVDLFR--NGDTSLSTIENYISAQAIVQGISNPSGDLSSGAGLGEPEKFNVDERTAYT 141
DB 223 NVLVEYENNTLSGNKKTILNVLKDYVTPSVKTSYSTVCN-----CGGEKFPNDSGYR 276
QY 142 GSGMRFPDRDPALRATMIGFG-OMLIDNGYTATDIWPLVRNDLSYVAQVWNOTGYD 200
DB 277 GAWGRPONDGPAAERATFTFLPADSYLTQTKDASVYVTLKPAIFKDLVYVNVWWSNCCFD 336
QY 201 LMEEVNGSSPFTTAVOHRALVEGSAFATVAGSS--GSCWDSQAPELICYLOSFWTGSFIL 258
DB 337 LMEEVNGVHFTYTLMWKRGKILGADPFKRGDSIRASTYSTSTINKTSSFWVSSNMW 336
QY 259 ANPDS-----SRSGKANTL---LQSIHTFPPEAACDSTFQPCSPRALANKEVVD 308
DB 397 IQVSQSVTGVSVKKGGLVSTLTAANLGSV-----DGGFTPGESEKTLATVANEDS 447
QY 309 FRSIYTLNDGLSDEAFAVAGRYPEDTY-----YNGNFWPLCTIAAEQLYDALYQMDKQG 363

```

```

DB 448 FASLYPINKNLBSYLGNSIGRYPEDTYNGNGNSQGSWFLAVTGYALYTRAKKWIENG 507
QY 364 SLEETDVSIDPFKALYSDAATG-TYSSSSSTSSIVDAVTFADGFSIVETHAAGSGM 422
DB 508 GVTYSSISLSPFFKKPSSSATSGKKYTVGTSDFNNLALQNLALADRSLSTYQJLHANNGL 567
QY 423 SEQYKSDGEBQSLARDLTWSYAALLTAN 450
DB 568 AEEFDRRTGSLSTGARDLTWSSHASLITAS 595
RESULT 40
AAR77674
ID AAR77674 standard; protein; 624 AA.
XX
XX AAR77674;
AC
XX 25-MAR-2003 (revised)
DT 29-JUN-1996 (first entry)
XX
XX Glucoamylase from Arxula adeninivorans.
DE
XX
XX glucoamylase; yeast; Saccharomyces cerevisiae; production; starch;
XX glucose; degradation; carbon source.
XX
XX Arxula adeninivorans.
OS
FH Key Location/Qualifiers
FT Peptide 1..16
FT /label= signal_peptide
FT Protein 17..624
FT /note= "for secretion of glucoamylase"
FT Modified-site 54..56
FT /note= "potential N-glycosylation site"
FT Modified-site 70..72
FT /note= "potential N-glycosylation site"
FT Modified-site 98..100
FT /note= "potential N-glycosylation site"
FT Modified-site 111..113
FT /note= "potential N-glycosylation site"
FT Modified-site 168..170
FT /note= "potential N-glycosylation site"
FT Modified-site 267..269
FT /note= "potential N-glycosylation site"
FT Modified-site 333..335
FT /note= "potential N-glycosylation site"
FT Modified-site 460..462
FT /note= "potential N-glycosylation site"
FT Modified-site /note= "potential N-glycosylation site"
PN DE4425058-AI.
XX
XX 18-JAN-1996.
PD
XX
XX 15-JUL-1994; 94DE-04425058.
PF
XX
XX 15-JUL-1994; 94DE-04425058.
PR
XX
XX (PFLA-) INST PFLANZENGENETIK & KULTURPFLANZENFOR.
PA
XX
XX Kunze G, Bui M, Kunze I, Foerster S;
XX
XX WPI, 1996-069579/08.
DR N-PSDB; AAT08701.
XX
XX
XX Prodn. of heat stable glucoamylase in yeast able to utilise starch - by
XX transformation with glucoamylase gene including secretion sequence from
XX Arxula adeninivorans.
XX
XX Claim 1; Fig 7; 21pp; German.
XX
XX The present sequence is a heat stable glucoamylase derived from Arxula
XX

```





```

XX XX MO8603778-A.
XX PN
XX 03-JUL-1986.
XX
XX PF 23-DEC-1985; 85NO-GB000599.
XX PR 21-DEC-1984; 84GB-00032483.
XX PR 21-DEC-1984; 86GB-00019568.
XX
XX PA (BREW-) BREWING RES FOUND.
XX PA (TUBB/) TUBB R. S.
XX PA (CLLT ) CELUTECH LTD.
XX
XX PI Tubb RS;
XX
XX WPI: 1986-182910/28.
XX DR N-PSDB; AAN60684.
XX
XX PT New precursor polypeptide of defined sequence - and corresp. DNA used to
XX transform hosts for prodn. of the polypeptide.
XX
XX PS Disclosure; Fig 4; 60pp; English.
XX
XX CC The inventors claim a DNA sequence encoding AAP60723 linked to a promoter
XX upstream and a gene for a polypeptide downstream. Particular examples are
XX the yeast enzyme AMG, the mammalian enzyme, gastric lipase and the
XX mammalian lymphokine, interferon-alpha2. 'X' in SQ indicates where a stop
XX codon in AAN60684 was translated. (Updated on 25-MAR-2003 to correct PA
XX field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
XX SQ Sequence 918 AA;

Query Match 21.7%; Score 600.5; DB 1; Length 918;
Best Local Similarity 29.6%; Pred. No. 4.5e-39;
Matches 164; Conservative 97; Mismatches 230; Indels 63; Gaps 17;

QY 1 MSFRLIALSGVCTGLANVISKRAITLDSWLSNEATVATKAILNIGADGAWVSGADSGI 60
DB 354 VQLRADVLMNGTVYD-SNGAMDSALAEWLQROKKVSIERIPENIGPSALYPS-ISP 411
QY 61 VVASPSTNDPNDYTYTWTROSGVLKTLVDLFRNGDTSLSTIENYISAQAIYVQGISNPSG 120
DB 412 VIASPSQTHDHYFYQWIRDSALITINSIVS--HSADPA-IEFLQYLVNSFHLQRTNNTLG 468
QY 121 D----LSSGAGLGEPEFVDETAATGSGRPPQBDPALRATAMIGFQMLDNG----- 170
DB 469 AGIGYNDYVALGDPKKNVNDTAFTBPGRRQNDGPAIRSAIILKIIDIYKQSGTDGAK 528
QY 171 --YTSTATDIWPLVRNDLSYVAQYNNQGYDLMEEVNGSSFITIAVQHRALVEGSAFAT 228
DB 529 YPFQSTF-DIFDDIVRWDLRFIIDHMNMSGFDLMEEVNGHFFTLVQLSAVDRLSYFN 587
QY 229 AVSSGSCMCD--SQAPEILCYL----QSFMTGSFILANF-----DSSRGSKDANTL 273
DB 588 ASERSSPEVELRQTRRDISKFLVDPANGFINGKY--NYIVETPMIADTLRSLDIDLSTL 644
QY 274 LGSIHFPDPAACDSTFOPCSPALANHKEVNDPSFSIYTLNDGLDSBAVAVGRYPED 333
DB 645 LAANTVHDAISR-SHLPFDINDPAVLNTLHMLHMSIYFINDSSGNAIGIALGRYPED 703
QY 334 TY-----YNGNPWELCTLAAAEOLYDALYQWQKQSLVTDVSLDFPKALYSDAATGYS 388
DB 704 VYDGYGVGEENPWLATCAASTLTYQLIYRHISQHDLVPMNNDCSNAFWSLVLFSNLT 763
QY 389 S-----SSSTYSSIVDAVKTFADGFSIVETHAASNSMSEQYDKSGEOLSA 436
DB 764 TLGNDGELYLLFETPTPAFNQTIQKIFQLADSFLVKAHAHGTDELSEQFNKRYTGFOGA 823
QY 437 RDLTWSAALITANRRNSVPA--SWGETSASSVGTCAA--TSAIGTY---SSVTV 487
DB 824 QHILWSTYSFMDAVQIQEVLQSLXTKKRKAASIHKICISXITNTIYTYIQNSDIT 883

```

```

QY 488 TSWPSIVATGCTT 501
DB 884 NXLDNVTGTSSPT 897

RESULT 43
AAP70183
ID AAP70183 standard; protein; 293 AA.
XX
XX AC AAP70183;
XX
XX DT 24-OCT-2003 (revised)
XX DT 25-MAR-2003 (revised)
XX DT 27-MAY-1991 (first entry)
XX
XX DE Sequence encoded by signal sequence of S.diastaticus DEX4 glucoamylase
XX gene.
XX
XX KM Yeast expression vector.
XX
XX OS Saccharomyces cerevisiae var. diastaticus.
XX
XX PN EP228254-A.
XX
XX PD 08-JUL-1987.
XX
XX PF 18-DEC-1986; 86EP-00309904.
XX
XX PR 18-DEC-1985; 85US-00810423.
XX
XX PA (BIOY ) BIOTECHNICA INT INC.
XX
XX PI Maine GT, Daves RS, Yocum RR;
XX
XX DR WPI: 1987-186757/27.
XX DR N-PSDB; AAN70265.
XX
XX PT Vector for transforming yeast - having secretory signal-encoding sequence
XX of glucocamylase gene from Saccharomyces diastaticus or S-cerevisiae.
XX
XX PS Disclosure; Fig 3; 11pp; English.
XX
XX CC AAN70265 contains the preferred signal sequence of the claimed vector.
XX CC The vector can be used to transform yeast cells to produce a desired
XX CC protein, eg glucocamylase from A.niger which can be used in brewing, corn
XX CC syrup prodn. and grain fermentation for distd. EtOH. prodn. (Updated on
XX CC 25-MAR-2003 to correct PA field.) (Updated on 24-OCT-2003 to standardise
XX CC OS field)
XX
XX SQ Sequence 293 AA;

Query Match 13.5%; Score 374; DB 1; Length 293;
Best Local Similarity 35.2%; Pred. No. 1.5e-21;
Matches 86; Conservative 45; Mismatches 89; Indels 24; Gaps 7;

QY 1 MSFRLIALSGVCTGLANVISKRAITLDSWLSNEATVATKAILNIGADGAWVSGADSGI 60
DB 56 VQLRADVLMNGTVYD-SNGAMDSALAEWLQROKKVSIERIPENIGPSALYPS-ISP 113
QY 61 VVASPSTNDPNDYTYTWTROSGVLKTLVDLFRNGDTSLSTIENYISAQAIYVQGISNPSG 120
DB 114 VIASPSQTHDHYFYQWIRDSALITINSIVS--HSADPALETLLQYLVNSFHLQRTNNTLG 167
QY 121 DLSGAG-----LGEPEFVDETAATGSGRPPQBDPALRATAMIGFQMLDNG--- 170
DB 168 TLGAGIGYNDYVALGDPKKNVNDTAFTBPGRRQNDGPAIRSAIILKIIDIYKQSGTDL 227
QY 171 ----YTSTATDIWPLVRNDLSYVAQYNNQGYDLMEEVNGSSFITIAVQHRALVEGSA 225
DB 228 GAKYPFQSTF-DIFDDIVRWDLRFIIDHMNMSGFDLMEEVNGHFFTLVQLSAVDRLSKLS 286
QY 226 FATA 229

```

DB 287 YFNA 290

RESULT 44  
ADC01413  
ID ADC01413 standard; protein; 1588 AA.

AC ADC01413;  
XX  
XX 04-DEC-2003 (first entry)  
DT  
XX Enterohaemorrhagic E. coli O157:H7-specific protein SEQ ID NO: 1459.  
DE  
XX enterohaemorrhagic; anti-bacterial.  
KM  
XX Escherichia coli; O157:H7.  
OS  
XX JP2002355074-A.  
PN  
XX 10-DEC-2002.  
PD  
XX 24-JAN-2002; 2002JP-00015959.  
PF  
XX 24-JAN-2001; 2001JP-00112010.  
PR  
XX (UYTS-) UNIV TSUKUBA.  
PA  
XX WPI; 2003-451640/43.  
DR  
XX  
XX Enterohaemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule  
PT and a polypeptide and its use, a polypeptide, a vector and a host cell.  
PS  
XX Claim 3; SEQ ID NO 1459; 2067pp; Japanese.

CC The invention relates to a novel enterohaemorrhagic Escherichia coli  
CC O157:H7-specific nucleic acid molecule. A polynucleotide of the invention  
CC has anti-bacterial activity. The polypeptide can be used in detection  
CC and/or treatment of O157:H7 infection. The nucleotide sequence of the  
CC genome of Enterohaemorrhagic E. coli O157:H7 was determined. The present  
CC sequence represents an E. coli O157:H7-specific polypeptide of the  
CC invention.  
CC  
SQ Sequence 1588 AA;

Query Match 6.5%; Score 181; DB 7; Length 1588;  
Best Local Similarity 21.7%; Pred. No. 7.2e-05; Indels 178; Gaps 30;  
Matches 136; Conservative 95; Mismatches 217;

QY 15 TGLANVISKRATLDSWLSNEATVARTAILNNIGADG-AMVSGADSGIVASPTDNDPDPF 73  
DB 576 TNLANTSTNATTTNINISN-----LTETVNLGEDALKM--DKDNGVFTAAHGETTETSKI 628  
QY 74 YWTRD-----SGLVLTLYDLFENGDTSLSTIENYISAQAIVQGISNPSGDL 122  
DB 629 -FNVKGDGLTTSSTDAVNGSQLKTNDVAVNTTNTATNTN-----ISNLTETV 677  
QY 123 SSAGAGGER-KNVDETATVSGWRPQRDGPALRATAMGRCOMLIDNYTATDIWM 180  
DB 678 TN---GGEALAKWKDKNGVFTAAHG---NNTASKITN-----ILDGVTATSSDAIN 723  
QY 181 PLVRNRL-SYVAQYM-----NOTG-----YDLMEEVNSSFPTIAVQHRALVESGAPA 227  
DB 724 GSQLYLLSSNIATYFGNANSVNTDGVFTGPTTKIGF-----TNNYNGDALAAI--NSSFS 777  
QY 228 TAVSSSCSWCDQAPEILCYLQSPWTSFI--LANPDSRSKGDG--NTLLGSIHTF--- 280  
DB 778 TSLGDLALW-DATAGKFSAGKGTNGDASVITVDADGHSIDSSSDAVAVNGSLHGVSSVVD 836  
QY 281 ----DEACDDSTFQPCSPRALANKKEVVDSEFRST-YTINDGSDSEAVAVGRYEDTY 335  
DB 837 ALGGGAEVNADGTTTAPTITIANADYDNVGDALNALIDTTLDDAL-----LMDADAG 887  
QY 336 YNGNPWFLLCTLAAEQLYDALYQMDKQGLEVTDVSLDFKALYSDAATG----- 185

DB 888 ENG-----AFSAHGXKOKTASV-ITNVANCAISAASDAINGSQLYTTNKYI 933  
QY 386 -----TYSSSSSTYSIYDAVTF-----A 405  
DB 934 ADALGGDAEVNADGTTTAPTITIANAEYNNVGDALDALLMDDETANCGAGAVNASH 993  
QY 406 DGFVSIVETHAASNGSMSE-QYDKSPDEQLSA-----RDLTWSYAL 446  
DB 994 DKAASITN--VANGSISEDSTDAVNGSQLMNTMMIEQNTQIINQLAGTDAITYOENG 1051  
QY 447 LTAANNRNSVWPASWETSASSVPGTCAATSAIGTSSVTVTSWPSIVATGTTTATPT 506  
DB 1052 AGINIVRTNDDGLAFNDASAQGVGAT-----AIG-YNSVAKGDSVAIGGYSVDVTGI 1105  
QY 507 GSGSVTSTSKTTRTATSKISTTTRSGM 532  
DB 1106 ALGSSSVSSSRVIAKGRDTSITENG 1131

RESULT 45  
AAR22343  
ID AAR22343 standard; protein; 32 AA.

AC AAR22343;  
XX  
XX 25-MAR-2003 (revised)  
DT 08-MAY-1992 (first entry)  
XX  
XX His(184) mutation in A. awamori glucoamylase region 4.  
DE  
XX 1,4-alpha-D-glucan glucosylhydrolase; EC 3.2.1.3; starch hydrolase.  
KM  
XX Aspergillus awamori.  
OS  
XX MO9200381-A.  
PN  
XX 09-JAN-1992.  
PD  
XX 29-JUN-1990; 90US-00546511.  
PF  
XX 29-JUN-1990; 90US-00546511.  
PR  
XX 29-JUN-1990; 90US-00546511.  
PA  
XX (NOVO) NOVO-NORDISK AS.  
PI  
XX Svensson KB, Sierks MR;  
XX  
XX N-PSDB; AAQ20804.  
DR  
XX  
XX New mutated glucoamylase enzymes - having aminoacid changes for increased  
PT selectivity for alpha-(1,4)-glucoside bonds in starch hydrolysis.  
PS  
XX Claim 18; Fig 2; 42pp; English.

CC This sequence represents amino acids 165 to 196 of the A. awamori  
CC glucoamylase enzyme, but with Ser 184 mutated to His. The mutation is in  
CC the region corresponding to Region 4 (residues 172-184) of the A. niger  
CC enzyme. The mutation confers increased selectivity for maltose hydrolysis  
CC on the enzyme expressed in Saccharomyces cerevisiae. (Updated on 25-MAR-  
CC 2003 to correct PA field.)  
CC  
SQ Sequence 32 AA;

Query Match 6.2%; Score 172; DB 2; Length 32;  
Best Local Similarity 96.9%; Pred. No. 9.9e-07;  
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 189 YVAQYNNQGYDLMEEVNSSFPTIAVQHRAL 220  
DB 1 YVAQYNNQGYDLMEEVNSSFPTIAVQHRAL 32

Mon Jun 28 08:40:17 2004

us-10-038-723-2.rag

Page 32

Search completed: June 28, 2004, 07:43:20  
Job time : 65 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 07:40:37 ; Search time 46 Seconds  
(without alignments)  
3662.755 Million cell updates/sec

Title: US-10-038-723-2

Perfect score: 2771

Sequence: 1 MSFRSLALSGLVCTGLANV.....SKTTATASKSTTTTSGMSL 534

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 100 summaries

Database :

SPTRMBL.25:\*

- 1: sp archea:\*
- 2: sp bacteria:\*
- 3: sp fungi:\*
- 4: sp human:\*
- 5: sp\_invertebrate:\*
- 6: sp mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2669.5	96.3	639	3	Q870G8 aspergillus
2	2564.5	92.5	639	3	Q12537 aspergillus
3	1689	61.0	618	3	Q9C1V4 calarymyces
4	1518	54.8	493	3	Q59846 aspergillus
5	1400	50.5	620	3	Q12523 humicola gr
6	1394	50.3	579	3	Q12596 corticium r
7	1286	46.4	571	3	Q9P4C5 lentiuma e
8	784	28.3	609	3	Q8J0P8 mucor circi
9	728	26.3	579	3	Q727X9 rhizopus or
10	674	24.3	515	3	Q8TFE5 saccharomyc
11	187	6.7	589	17	Q96Z10 sulfolobus
12	185.5	6.7	436	17	Q9H188 thermoplaam
13	181	6.5	1588	16	Q8XDG4 escherichia
14	175	6.3	1236	3	Q9C105 schizosacch
15	174.5	6.3	659	17	Q97BM7 thermoplaam
16	174.5	6.3	990	16	Q83J22 shigella fl

17	173.5	6.3	1713	3	Q8TGE1	Q8TGE1 saccharomyc
18	172	6.2	622	17	Q97ZD0	Q97ZD0 sulfolobus
19	171	6.2	3360	16	Q88XB6	Q88XB6 lactobacill
20	170.5	6.2	1275	5	Q76602	Q76602 caenorhabdi
21	169.5	6.1	190	3	Q07070	Q07070 saccharomyc
22	168	6.1	4106	16	Q8XQP2	Q8XQP2 raietonia s
23	164	5.9	2016	5	Q9B1T0	Q9B1T0 plectureuys
24	163.5	5.9	457	5	Q86AK1	Q86AK1 dictyosteli
25	163	5.9	1333	16	Q8PD38	Q8PD38 xanthomonas
26	161	5.8	1063	16	Q8CH86	Q8CH86 lactococcus
27	161	5.8	2283	2	Q8VQ99	Q8VQ99 staphylococ
28	160.5	5.8	916	5	Q7YZ10	Q7YZ10 monosiga br
29	160	5.8	615	17	Q973T2	Q973T2 sulfolobus
30	160	5.8	888	5	Q25336	Q25336 leishmania
31	159	5.7	2271	16	Q99QV4	Q99QV4 staphylococ
32	158	5.7	1283	3	Q9URU4	Q9URU4 schizosacch
33	156.5	5.6	2232	5	Q81FX6	Q81FX6 caenorhabdi
34	156	5.6	1306	2	Q93N36	Q93N36 pantoea ana
35	154.5	5.6	2275	16	Q8NUJ3	Q8NUJ3 staphylococ
36	153	5.5	1131	3	Q74851	Q74851 schizosacch
37	153	5.5	2117	16	Q8B473	Q8B473 streptococc
38	152.5	5.5	2117	3	Q96U11	Q96U11 neurospora
39	152	5.5	2014	16	Q7U3X4	Q7U3X4 synecococc
40	150	5.4	1383	3	Q874K9	Q874K9 candida alb
41	149	5.4	967	3	Q08294	Q08294 saccharomyc
42	149	5.4	1001	3	Q05164	Q05164 saccharomyc
43	147	5.3	2230	16	Q7U7J7	Q7U7J7 synecococc
44	146	5.3	612	17	Q97VY0	Q97VY0 sulfolobus
45	145	5.2	860	16	Q88T89	Q88T89 lactobacill
46	144.5	5.2	614	16	Q9CHH3	Q9CHH3 lactococcus
47	144	5.2	948	3	Q7A346	Q7A346 schizosacch
48	143.5	5.2	881	2	Q9AF09	Q9AF09 cellvibrrio
49	143	5.2	1995	17	Q8T164	Q8T164 methanosarc
50	142	5.1	536	3	Q9UVS8	Q9UVS8 aspergillus
51	142	5.1	854	16	Q89GQ1	Q89GQ1 bradyrhizob
52	142	5.1	1672	16	Q8Y366	Q8Y366 raietonia s
53	141.5	5.1	1441	16	Q9CF11	Q9CF11 lactococcus
54	141.5	5.1	3145	16	Q98MG7	Q98MG7 rhizobium l
55	140.5	5.1	1055	16	Q839R5	Q839R5 enterococcu
56	140.5	5.1	1952	5	Q8WVW6	Q8WVW6 naegleria g
57	140.5	5.1	4776	16	Q97P71	Q97P71 streptococc
58	140.5	5.1	5188	16	Q8X4H5	Q8X4H5 escherichia
59	140	5.1	1802	3	Q04051	Q04051 saccharomyc
60	140	5.1	1994	16	Q8G521	Q8G521 bifidobacte
61	139.5	5.0	2522	16	Q8EKA6	Q8EKA6 shewanella
62	139.5	5.0	4654	16	Q8D418	Q8D418 vibrio vuln
63	139	5.0	402	10	Q23054	Q23054 arabidopsia
64	139	5.0	744	3	Q8TFG9	Q8TFG9 schizosacch
65	139	5.0	970	16	Q8DYL7	Q8DYL7 streptococc
66	139	5.0	1203	5	Q9N5K0	Q9N5K0 caenorhabdi
67	139	5.0	22152	4	Q8W417	Q8W417 homo sapien
68	138.5	5.0	600	5	Q86B01	Q86B01 dictyosteli
69	138.5	5.0	1589	5	Q46097	Q46097 drosophila
70	138.5	5.0	2310	16	Q8CMU7	Q8CMU7 staphylococ
71	138.5	5.0	5291	16	Q8X2T1	Q8X2T1 escherichia
72	138	5.0	316	12	Q8VOM4	Q8VOM4 equine herp
73	138	5.0	332	12	Q8VOM5	Q8VOM5 equine herp
74	138	5.0	337	12	Q8VOM1	Q8VOM1 equine herp
75	138	5.0	342	12	Q8VOM8	Q8VOM8 equine herp
76	138	5.0	356	12	Q8VOM7	Q8VOM7 equine herp
77	138	5.0	357	12	Q8VOM2	Q8VOM2 equine herp
78	138	5.0	372	12	Q8VOM3	Q8VOM3 equine herp
79	138	5.0	374	12	Q8VOM6	Q8VOM6 equine herp
80	138	5.0	389	12	Q8VOM0	Q8VOM0 equine herp
81	138	5.0	826	12	Q8VOM5	Q8VOM5 equine herp
82	138	5.0	867	12	Q39782	Q39782 equine herp
83	138	5.0	1347	3	Q874L0	Q874L0 candida alb
84	138	5.0	1778	16	Q8RCB2	Q8RCB2 escherichia
85	137.5	5.0	481	5	Q9XUF4	Q9XUF4 caenorhabdi
86	137.5	5.0	625	3	Q74922	Q74922 schizosacch
87	137	4.9	354	16	Q88TS6	Q88TS6 lactobacill
88	136.5	4.9	1265	16	Q8EV72	Q8EV72 mycoplasma
89	136	4.9	314	16	Q88TS7	Q88TS7 lactobacill

90 136 4.9 866 12 039781 039781 equine herp  
91 136 4.9 1180 16 09CHM4 09CHM4 lactococcus  
92 136 4.9 3072 2 0939M5 0939M5 streptococc  
93 135.5 4.9 496 16 088VA6 088VA6 lactobacill  
94 135.5 4.9 658 17 08PYJ7 08PYJ7 methanocarc  
95 135 4.9 699 2 048494 048494 kurtzia zop  
96 135 4.9 3624 16 082411 082411 salmonella  
97 134.5 4.9 651 2 09KMR2 09KMR2 thermocactin  
98 134.5 4.9 930 2 09RFK5 09RFK5 caldibacill  
99 134.5 4.9 1014 5 0962B4 0962B4 helicoverpa  
100 134 4.8 498 5 09GOU3 09GOU3 cheilleria p

## ALIGNMENTS

RESULT 1  
0870G8 PRELIMINARY; PRT; 639 AA.  
ID 0870G8  
AC 0870G8  
DT 01-JUN-2003 (TEMBLrel. 24, Created)  
DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
DE Glucoamylase  
OS Aspergillus niger.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eukaryotes; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5061;  
[1]  
RN RP SEQUENCE FROM N.A.  
RC STRAIN=721;  
RX MEDLINE=95066018; PubMed=7975554;  
RA Zhong L., Tang G., Yang K.;  
RT "Isolation and sequencing of glucoamylase gene from a glucoamylase  
RT over producing strain."  
RL Wei Sheng Wu Xue Bao 34:184-190(1994).  
[2]  
RN RP SEQUENCE FROM N.A.  
RC STRAIN=721;  
RX MEDLINE=95066018; PubMed=7975554;  
RA Zhong L., Tang G., Yang K.;  
RT "Cloning, sequencing and comparison of the 5' flanking regions of glaa  
RT gene from high and low glucoamylase-producing strains of Aspergillus  
RT niger."  
RL Wei Sheng Wu Xue Bao 36:181-186(1996).  
[3]  
RN RP SEQUENCE FROM N.A.  
RC STRAIN=721;  
RA Zhong L., Qiao D., Tang G., Yang K.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY505996; AAD04499.1; -  
DR GO; GO:0003824; F:catalytic activity; IEA.  
DR GO; GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.  
DR GO; GO:0005976; P:polysaccharide metabolism; IEA.  
DR InterPro; IPR008291; Glu-a-glcSD\_SBD.  
DR InterPro; IPR008291; Glu-a-glcSD\_SBD.  
DR InterPro; IPR000165; Glyco\_hydro\_15.  
DR InterPro; IPR008928; Glyco\_trans\_6np.  
DR Pfam; PF00686; CBM\_20; 1.  
DR Pfam; PF00723; Glyco\_hydro\_15; 1.  
DR PRINTS; PR00736; GLHYDRLASE15.  
DR ProDom; PD001568; CBD\_4; 1.  
DR PIRSF; PIRSF001031; Glu-a-glcSD\_SBD; 1.  
DR PIRSF; PIRSF001031; Glu-a-glcSD\_SBD; 1.  
SQ SEQUENCE 639 AA; 68308 MW; 025AD856B6542B69 CRC64;

Query Match 96.3%; Score 2669.5; DB 3; Length 639;  
Best Local Similarity 98.1%; Pred. No. 3.9e-162;  
Matches 521; Conservative 1; Mismatches 6; Indels 3; Gaps 3;

QY 1 MSFRSLALSGVCTGLANYSKRATLDSWLSNEATVARTALININIGADGAWVSGADSGI 60  
Db 1 MSFRSLALSGVCTGLANYSKRATLDSWLSNEATVARTALININIGADGAWVSGADSGI 60  
QY 61 VVASPSTDNPDYFTYTRDSGLVLTIVLFRNGDTSLSTIENYISAQALVQGISNPSG 120

Db 61 VVASPSTDNPDYFTYTRDSGLVLTIVLFRNGDTSLSTIENYISAQALVQGISNPSG 120  
QY 121 DLSSGAGLEPKFENVETAATGSGWRPORDPALRATAMIGFQWMLDNGYSTANDIYW 180  
Db 121 DLSSGAGLEPKFENVETAATGSGWRPORDPALRATAMIGFQWMLDNGYSTANDIYW 180  
QY 181 PLVRNDLSYVAQYWNQTDGLDEEVNCGSFETIYVQHRALVEGSAFATVAGSSCMWDSQ 240  
Db 181 PLVRNDLSYVAQYWNQTDGLDEEVNCGSFETIYVQHRALVEGSAFATVAGSSCMWDSQ 240  
QY 241 APEILCYLQSFMTGFTLANFDSRSRSGKANT-LDSIHFPDEAACDDSTFQPCSPRAL 299  
Db 241 APEILCYLQSFMTGFTLANFDSRSRSGKANT-LDSIHFPDEAACDDSTFQPCSPRAL 299  
QY 300 ANKEVVDSEFRSIYTLNDGLSDSEAVAVGRYPEDTYNCPWFLCTLAALAEQYDALYOW 359  
Db 300 ANKEVVDSEFRSIYTLNDGLSDSEAVAVGRYPEDTYNCPWFLCTLAALAEQYDALYOW 359  
QY 360 DKQSLAEVTDVSLDFPKALYSDAATGTYSSSSSTYSYDAVKTFPADGPIYSIVETHAASN 419  
Db 360 DKQSLAEVTDVSLDFPKALYSDAATGTYSSSSSTYSYDAVKTFPADGPIYSIVETHAASN 419  
QY 420 GSWSEQYDSDGQSLARDTWISYALALTANRRNSVVPASWGETSASVPGTCATSAI 479  
Db 420 GSWSEQYDSDGQSLARDTWISYALALTANRRNSVVPASWGETSASVPGTCATSAI 479  
QY 480 GTYSVTVTSWPSIVATGTTTATPTGSGSVTSKTTATASKTSTTRS 530  
Db 480 GTYSVTVTSWPSIVATGTTTATPTGSGSVTSKTTATASKTSTTRS 530  
QY 479 GTYSVTVTSWPSIVATGTTTATPTGSGSVTSKTTATASKTSTTRS 529  
Db 479 GTYSVTVTSWPSIVATGTTTATPTGSGSVTSKTTATASKTSTTRS 529

RESULT 2  
012537 PRELIMINARY; PRT; 639 AA.  
ID 012537  
AC 012537  
DT 01-NOV-1996 (TEMBLrel. 01, Created)  
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
DE Glucoamylase precursor (EC 3.2.1.3).  
GN GLA.  
OS Aspergillus awamori.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eukaryotes; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=105351;  
[1]  
RN RP SEQUENCE FROM N.A.  
RC STRAIN=X-100;  
RA Diagne I., Pechexonov V.T., Bulat S.A., Firov L.M.;  
RT "A glucoamylase gene from Aspergillus awamori X-100: structure,  
RT allocation, and gene phylogeny."  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U59303; AAB02927.1; -  
DR HSSP; P04064; IGAT.  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR GO; GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.  
DR GO; GO:0016798; F:hydrolase activity; acting on glycosyl bonds; IEA.  
DR GO; GO:0005976; P:polysaccharide metabolism; IEA.  
DR InterPro; IPR002044; CBD\_4.  
DR InterPro; IPR002044; EF-Hand.  
DR InterPro; IPR008291; Glu-a-glcSD\_SBD.  
DR InterPro; IPR000165; Glyco\_hydro\_15.  
DR InterPro; IPR008928; Glyco\_trans\_6np.  
DR Pfam; PF00686; CBM\_20; 1.  
DR Pfam; PF00723; Glyco\_hydro\_15; 1.  
DR PRINTS; PR00736; GLHYDRLASE15.  
DR ProDom; PD001568; CBD\_4; 1.  
DR PROSITE; PS00018; EF\_HAND; 1.  
DR PROSITE; PS00820; GLUCOAMYLASE; 1.  
DR PIRSF; PIRSF001031; Glu-a-glcSD\_SBD; 1.  
KW Glycosidase; Hydrolase; Signal.  
FT SIGNAL 1 24  
FT CHAIN 25 639 GLUCOAMYLASE.

SQ SEQUENCE 639 AA; 68277 MW; 6F93D0637D174ACB CRC64;  
 Query Match 92.5%; Score 2564.5; DB 3; Length 639;  
 Best Local Similarity 92.3%; Pred. No. 2e-155; Mismatches 14; Indels 11; Gaps 3;  
 Matches 494; Conservative 16; Mismatches 14; Indels 11; Gaps 3;  
 QY 1 MSFRSLALSGVCTGLANVYISKRATLDSWLSNEATVARTAILNNICADGAWYSGADSGI 60  
 DB 1 MSFRSLALSGVCTGLANVYISKRATLDSWLSNEATVARTAILNNICADGAWYSGADSGI 60  
 QY 61 VVASPTNDPDYFYTWTRDSGLVLTVDLFRNGDTSLSTENYISAQIVGQISNPSG 120  
 DB 61 VVASPTNDPDYFYTWTRDSGLVLTVDLFRNGDTSLSTENYISAQIVGQISNPSG 120  
 QY 121 DLSGGGCEPKKNVDETAATGSGWRPQRDPALRTAMIGFQWMLDNGYSTATDIW 180  
 DB 121 DLSGGGCEPKKNVDETAATGSGWRPQRDPALRTAMIGFQWMLDNGYSTATDIW 180  
 QY 181 PLVRNDLSYVAQYMNQGYDLMEEVNGSFFFTIAYQHRALVEGSAFATVAGSSCSCWCDSDQ 240  
 DB 180 PLVRNDLSYVAQYMNQGYDLMEEVNGSFFFTIAYQHRALVEGSAFATVAGSSCSCWCDSDQ 240  
 QY 241 APEILCYLOSF-----WTGSFIILNFDSSRSGKQANTLGSHTFPEAACDDSTFQPCS 295  
 DB 240 APEILCYLOSF-----WTGSFIILNFDSSRSGKQANTLGSHTFPEAACDDSTFQPCS 295  
 QY 296 PRLANHKEVVDSPRSTIYTLNDLSSEAVAVGRYPEDTYNNGNPFELCTLAAEQLYDA 355  
 DB 295 PRLANHKEVVDSPRSTIYTLNDLSSEAVAVGRYPEDTYNNGNPFELCTLAAEQLYDA 355  
 QY 356 LVQMDKQSGLEVVDLDFPKALYSDAATGTYSSSTYSIVDAVTFADGFSIVETH 415  
 DB 355 LVQMDKQSGLEVVDLDFPKALYSDAATGTYSSSTYSIVDAVTFADGFSIVETH 415  
 QY 416 AASNGSSEQYDKSDDEQLSARDLTWSYALLTANNRRNSVPAWGETSASSVPGTCAA 475  
 DB 415 AASNGSSEQYDKSDDEQLSARDLTWSYALLTANNRRNSVPAWGETSASSVPGTCAA 475  
 QY 476 TSAIGTYSSVTVSWPSIVATGTTTATPTGSGSVTSKTTATASKSTTTRS 530  
 DB 475 TSAIGTYSSVTVSWPSIVATGTTTATPTGSGSVTSKTTATASKSTTTRS 529  
 RESULT 3  
 QY 09C1V4 PRELIMINARY; PRT; 618 AA.  
 AC 09C1V4;  
 DT 01-JUN-2001 (TEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE Glucanase precursor (EC 3.2.1.3).  
 GN GA.  
 OS Talaromyces emersonii.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiiales; Trichocomaceae; Talaromyces.  
 OX NCBI\_TaxID=68825;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nielsen B.R., Lehmbeck J., Frandsen T.P.;  
 RT "Cloning, heterologous expression, and enzymatic characterization of a  
 thermostable glucanase from Talaromyces emersonii.";  
 RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AJ304803; CAC28076.1; -.  
 DR HSSP; P04064; IGAT.  
 DR GO; GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.  
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
 DR GO; GO:0005976; P:polysaccharide metabolism; IEA.  
 DR InterPro; IPR008291; Glu-a-glcSD\_SBD.  
 DR InterPro; IPR000165; Glyco\_hydro\_15.  
 DR InterPro; IPR008928; Glyco\_trans\_6np.  
 DR Pfam; PF00686; CBM\_20; 1.  
 DR Pfam; PF00723; Glyco\_hydro\_15; 1.

DR PRINTS; PR00736; GLHYDRASE15.  
 DR ProDom; PD001568; CBD\_4; 1.  
 DR PIRSF; PIRSF001031; Glu-a-glcSD\_SBD; 1.  
 KM Glycosidase; Hydrolase; Signal.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 28 618 GLUCANYLASE.  
 SQ SEQUENCE 618 AA; 65429 MW; C48A034A2C06B5B2 CRC64;  
 Query Match 61.0%; Score 1689; DB 3; Length 618;  
 Best Local Similarity 61.2%; Pred. No. 1.4e-99; Mismatches 111; Indels 18; Gaps 5;  
 Matches 322; Conservative 75; Mismatches 111; Indels 18; Gaps 5;  
 QY 7 LALSGVCTGLANV-ISKAT--LDSWLSNEATVARTAILNNICADGAWYSGADSGI 63  
 DB 9 LCLIGTPAFAAPAAATASGLDSFLATETPILALQVLNNIGPAGADVAGSAGI 68  
 QY 64 SPSTNDPDYFYTWTRDSGLVLTVDLFRNGDTSLSTENYISAQIVGQISNPSG 123  
 DB 69 SPSTNDPDYFYTWTRDSGLVLTVDLFRNGDTSLSTENYISAQIVGQISNPSG 128  
 QY 124 SGAGIGEPKKNVDETAATGSGWRPQRDPALRTAMIGFQWMLDNGYSTATDIW 183  
 DB 129 TG-GLGEPKKNVDETAATGSGWRPQRDPALRTAMIGFQWMLDNGYSTATDIW 187  
 QY 184 RNDLSYVAQYMNQGYDLMEEVNGSFFFTIAYQHRALVEGSAFATVAGSSCSCWCDSDQ 243  
 DB 188 QNDLSITTYQYMNSTDLMEEVNGSFFFTIAYQHRALVEGSAFATVAGSSCSCWCDSDQ 247  
 QY 244 ILCYLOSFMTGFIILNFDSSRSGKQANTLGSHTFPEAACDDSTFQPCS 302  
 DB 248 VLFQFQSVYTNVLFNGSGRSGKQANTLGSHTFPEAACDDSTFQPCS 307  
 QY 303 KEVVDSPRSTIYTLNDLSSEAVAVGRYPEDTYNNGNPFELCTLAAEQLYDA 362  
 DB 308 KEVVDSPRSTIYTLNDLSSEAVAVGRYPEDTYNNGNPFELCTLAAEQLYDA 367  
 QY 363 GSELEVVDLDFPKALYSDAATGTYSSSTYSIVDAVTFADGFSIVETHAASNGSM 422  
 DB 368 GSISITVDLDFPKALYSDAATGTYSSSTYSIVDAVTFADGFSIVETHAASNGSM 427  
 QY 423 SEQYDKSDDEQLSARDLTWSYALLTANNRRNSVPAWGETSASSVPGTCAATSAIGTY 482  
 DB 428 TEQFSRTDGPASALVTSYASLTASARQGVPAWGESASSVPAWCAATSAIGTY 487  
 QY 483 SSVTVSWPSIVATGTTTATPTGSGSVTSKTTATASKSTT 528  
 DB 488 STATNTWPS-----SGSGSSTTSSAPCTTPTSAVAT 520  
 RESULT 4  
 QY 059846 PRELIMINARY; PRT; 493 AA.  
 AC 059846;  
 DT 01-AUG-1998 (TEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE Glucanase.  
 OS Aspergillus oryzae.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_TaxID=5062;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O-1013;  
 RA MEDLINE=98172744; PubMed=9511753;  
 RA Hata Y., Ishida H., Ichikawa E., Kawato A., Suginami K., Imai S.,  
 RT "Nucleotide sequence of an alternative glucanase-encoding gene  
 (GlaB) expressed in solid-state culture of Aspergillus oryzae.";  
 RL Gene 207:127-134 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O-1013;  
 RA Hata Y., Ishida H., Kojima Y., Ichikawa E., Kawato A., Suginami K.,



RA Imayasu S.;  
RT "Comparison of two glucoamylases produced by *Aspergillus oryzae* in  
solid-state (koji) and in submerged culture.";  
RL J. Ferment. Biotech. 84:532-537(1997).  
DR EMBL: AB007825; BAA25205.1; -.  
DR PIR: PC6503; JG6538.  
DR HSSP: P04064; 1GAT.  
DR GO: GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.  
DR GO: GO:0005976; P:polysaccharide metabolism; IEA.  
DR InterPro: IPR00165; Glyco\_hydro\_15.  
DR InterPro: IPR008928; Glyco\_trans\_6hp.  
DR Pfam: PF00723; Glyco\_hydro\_15; 1.  
DR PRINTS: PR00736; GLHTRLASE15.  
DR PROSITE: PS00820; GLUCOAMYLASE; 1.  
DR SEQUENCE 493 AA; 52401 MW; 3707B3A6FC66256 CRC64;  
SQ  
Query Match 54.8%; Score 1518; DB 3; Length 493;  
Best Local Similarity 57.5%; Pred. No. 8.6e-89;  
Matches 280; Conservative 82; Mismatches 123; Indels 2; Gaps 2;  
QY 5 SLIALSGLVCTGLANVISKRATLDSMLNEATVARTAILNNIGADGAWGSGIVVAS 64  
DB 8 SLNIAIGAVAHPSPIHKRQSDLNAFIEAQTPIAKQGLVNNIGADGLVEGAAGIVVAS 67  
QY 65 PSTDNPDYFTTWRDGLVLTVDLFRNGDTSLSLTENTYISAQAIVQGISNPSGLSS 124  
DB 68 PLSKNPDYFTWTRDAGLTMEVEYEQPIGDDATLESTIQNYVDSQANEQAVSNPSGGLSD 127  
QY 125 GAGLGEKPFVNDATAYVSGRPGORDPALRATAMIGFGWLDNGVTATDITVWPLVR 184  
DB 128 GSGLAERKFTVNIQGFIDSMGRPGORDPALRASKMLVGNLSISDQSVVKANINPIVQ 187  
QY 185 NDLSYVAQYNNQGYDLMEEVNGSSFTTIAVORALVEGSAFATAVGSSGWCDSQAPET 244  
DB 188 NDLSTVQYNNQGYDLMEEVNGSSFTTIAVORALVEGSAFATAVGSSGWCDSQAPET 246  
QY 245 LCTVQSWTSSFLIANDSS-RSGKDANTLLGSHITDPPEAACDSTFFQCSPPALANK 303  
DB 247 LCHLQDPMNSAVLSNLPTRNGRGLDTPNSLLGSHITDPPEAACDSTFFQCSPPALANK 306  
QY 304 EYVDSFSLITLNDGLSDSEAVAGRYPEPTYNGNPMFLCTLAARALVDALYQMDKOG 363  
DB 307 LVVDSFSLVGINNGRAGKAAVGPYAEPTYGCGNMYLTLVAAELLYDALYQMDKOG 366  
QY 364 SLEVTVDLDFKALYSDAATGTYSSTYSIVDAVKTADGFSIVETHAASNGSMS 423  
DB 367 QVNVTEISLPEFKDLSNVTTGSAKSSAYESLTSVAKTYADGFISVQETPDGALA 426  
QY 424 EYDKSGEQLSARDLTWSTALILTANRRNSVVPASWGETSASVPGTCAITSIGTYS 483  
DB 427 EYKSRDQGTIVSADLTWSTAAFLSAVGRNRTVPASWGSTANAVPSCSGCTVSGSYT 486  
QY 484 SVTYTSM 490  
DB 487 TPTVGS# 493

RESULT 5  
ID 012623 PRELIMINARY; PRT; 620 AA.  
AC 012623;  
DT 01-NOV-1996 (TREMUREL. 01, Created)  
DT 01-NOV-1996 (TREMUREL. 01, Last sequence update)  
DT 01-OCT-2003 (TREMUREL. 25, Last annotation update)  
DE Glucoamylase (EC 3.2.1.3).  
GN GLA.  
OS Humicola grisea var. thermoides.  
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.  
OX NCB1\_TaxID=5528;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Berka R.M., Rey M.W., Thompson S.A., Gray G.L., Carmona C.L.,  
Power S.D.;

RT "Molecular cloning, analysis and expression of the glal gene encoding  
a thermostable, raw starch-digesting glucoamylase from the fungus  
Humicola grisea var. thermoides."  
RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
DR EMBL: M89475; AAA33366.1; -.  
DR HSSP: P04064; 1GAT.  
DR GO: GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.  
DR GO: GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
DR GO: GO:0005976; P:polysaccharide metabolism; IEA.  
DR InterPro: IPR002044; CBD 4.  
DR InterPro: IPR008291; Glu-a-glcad\_SBD.  
DR InterPro: IPR00165; Glyco\_hydro\_15.  
DR InterPro: IPR008928; Glyco\_trans\_6hp.  
DR Pfam: PF00686; CBM\_20; 1.  
DR Pfam: PF00723; Glyco\_hydro\_15; 1.  
DR Pfam: PD001568; CBD 4; 1.  
DR PROSITE: PS00820; GLUCOAMYLASE; 1.  
DR PROSITE: PS001031; Glu-a-glcad\_SBD; 1.  
DR Glycosidase; Hydrolase.  
DR SEQUENCE 620 AA; 66525 MW; 54214FF67E20B80A CRC64;  
SQ  
Query Match 50.5%; Score 1400; DB 3; Length 620;  
Best Local Similarity 49.5%; Pred. No. 4e-81;  
Matches 269; Conservative 96; Mismatches 146; Indels 32; Gaps 7;  
QY 2 SFRSLIATSGLVCTGL-----ANVISKRATLDSMLNEATVARTAILNNIGADGAWVSGA 56  
DB 3 TFEKLVLSGAVGSAIGRPHGSRRLQERAAVDFTNTEKRIANNKLLANIGPGRKAPGA 62  
QY 57 DSGIVVASPESTDNPDTFTTWRDGLVLTVDLFRNGDTSLSLTENTYISAQAIVQGI 115  
DB 63 AAGVVASPSRTPDPFFFTWTRDAGLTMEVEYEQPIGDDATLESTIQNYVDSQANEQAVSNPSGGLSD 108  
QY 116 SNPSGGLSSGAGLGEKPFVNDATAYVSGRPGORDPALRATAMIGFGWLDNGVTSTA 175  
DB 109 SNPSGTFPADSGGLEAKFVNDLTAFTEGMRPGORDPALRALIYQAKMLANGKSTK 168  
QY 176 TDIWPLVRNDLSYVAQYNNQGYDLMEEVNGSSFTTIAVORALVEGSAFATAVGSSCS 235  
DB 169 KSVVWPVVKNDLAVTAYQVNMETGFDLMEEVNGSSFTTIASSHRLAEGAVIAAQLDTECP 228  
QY 236 WCDSQAPETLCYQSWT--GSLTIANFDS--RSGKDANTLLGSHITDPPEAACDSTF 291  
DB 229 PCTTVAPQVYCFQOAFWNSKGNVSTTAGEYRSGKDANSILASIHNPDEAGCNLTTF 288  
QY 292 QPCSPPALANKHKNVSPRSITLNDGLSDSEAVAGRYPEPTYNGNPMFLCTLAARQ 351  
DB 289 QPCSEPPALANKHKNVSPRSITLNDGLSDSEAVAGRYSEDTYNGNPMFLANPAAEQ 348  
QY 352 LYDALYQMDKOGSLEVTVDLDFKALYSDAATGTYSSTYSIVDAVKTADGFSV1 411  
DB 349 LYDAIYVNNKQGISITTVSVSLPFRDLVSVSTGSKSSSTFTNIVNAKAAVDGTEV 408  
QY 412 VETHAASNGSMSQYDKSGEQLSARDLTWSTALILTANRRNSVVPASWGETSASV 470  
DB 409 AAKYTPSNGALAEQYDRNTGKPDAAADLTWSTAAFLSAVGRNRTVPASWGSTANAVP 468  
QY 471 GTCATSAITGYSVTVTWSPIVATGCTTTATPGSGSVTSKTTATASKTSTTRS 530  
DB 469 STCSRLEVAGTYVAAISTSPS-----KQTNPSAAPSPPYPPACADASEYVTFNE 521  
QY 531 GMS 533  
DB 522 RVS 524

RESULT 6  
ID 012596 PRELIMINARY; PRT; 579 AA.  
AC 012596;  
DT 01-NOV-1996 (TREMUREL. 01, Created)  
DT 01-NOV-1996 (TREMUREL. 01, Last sequence update)



```
QY 472 ---TCAATSAIGTY-----SSVTVTSWPSIVANGCTTTT---ATPFGSGVSTST- 515
Db 492 NVYLTGAVDALIEDWSTDNAILLSSANYPWSTVVDLPGSTDVQYKIKDGGSTVWESD 551
QY 516 ---KTATATKSTST 527
Db 552 PNMEITTPANGTYAT 566

RESULT 8
08J0P8
ID 08J0P8 PRELIMINARY; PRT; 609 AA.
AC 08J0P8;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Glucoamylase Glam.
OS Mucor circinellioidea.
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC Mucor.
OX NCBI_TaxID=36080;
RN [1]
RP SEQUENCE FROM N.A.
RA Houghton-Larsen J., Pedersen P.A.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY168303; AN85206.1; -
DR GO; GO:0004339; F:Glucan 1,4-alpha-glucosidase activity; IEA.
DR GO; GO:0005976; F:Polysaccharide metabolism; IEA.
DR InterPro; IPR005036; CBM_21.
DR InterPro; IPR000165; Glyco_hydro_15.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF03370; CBM 21; 1.
DR Pfam; PF00723; Glyco_hydro_15; 1.
DR PRINTS; PR00736; GLHYDOLASE15.
DR PROSITE; PS00820; GLUCOAMYLASE; 1.
SQ SEQUENCE 609 AA; 64778 MW; 56DD54CD371CC7C1 CRC64;

Query Match 28.3%; Score 784; DB 3; Length 609;
Best Local Similarity 39.1%; Pred. No. 8e-42;
Matches 175; Conservative 73; Mismatches 158; Indels 42; Gaps 11;

QY 25 ATLDWSLNSNATVARTAILNIGADGAWSGAGSIIVASPTDNPDTFTWTRDGLVL 84
Db 173 STSSWIDGQTSVRVAMLRNINP-----AGAVTGFLAASWSTGPDYFYAMTRDSALTS 227
QY 85 KTIYDLFRN---GDTSLSTIENYISAQAIYQGISNPSGDLSSGAGIGEPFNVDATYT 141
Db 228 HVAAYDYNNTLLAGNSTLLGLKNTVFTSLNSQTSVTCN-----CLGEPEFNKSSGYS 281
QY 142 GSWGRPQRDGPALRATAMIGFGQWL--DNGYTSTATDIWPLVHNDSLVAQYNNQTCYD 200
Db 282 GAWGRPQNDGPASADPFIILADSLIKQTDGATYVTGTLAPAIKLDYVSTSNCGFD 341
QY 201 LMEEVNGSSFFTLAVQHRALVEGSAFATAVGSS--CSWCDSQAPELICYLQSFN--TGSF 256
Db 342 LMEEVNGVHFYTLVWVRGVLKVGANFATRNDSFTRATYTTNTAASIKTKIDSFNNSNGQY 401
QY 257 ILAIFD-----SSRGKQANTL---LGSIHFPDEACDDSTPQPCSPRALANHKEVVD 308
Db 402 VVSQSVTGVSXKAGYASVLIAASNLGSL-----QDCFTPGSDKMLATAVAESK 452
QY 309 FRSIYTLNDGLDSEEAVALGRYPEDTY-----YNGNPFILCTLAAREQLYDALYQMDKOG 363
Db 453 FASIVSISNONINGVLGNAIGRYPEDTYNGNGNSQGNPFICTNAPALRYAALKEWFRNG 512
QY 364 SLEVTIVSLDFKALYSDATGT--YSSSSSTYSIIVDAVKTFFADGFSIVETHAASNGSM 422
Db 513 GVTYTSISLNFPPFKFSSAAAGTKYVGTSSFNLSLVQVAVAADAFSTIKFHAATNGSM 572
QY 423 SEQYKSDGDEQLSARDLTWGAALLTAN 450
Db 573 SEQYGRADGLMTGARDLTWASHASLISAS 600
```

```
RESULT 9
Q727X9
ID 0727X9 PRELIMINARY; PRT; 579 AA.
AC 0727X9;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Glucoamylase (Fragment).
OS Rhizopus oryzae (Rhizopus delemar).
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC Rhizopus.
OX NCBI_TaxID=64495;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen G., Yu X.C., Jiang H.Z., Li M.G.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY359821; AA018643.1; -.
FT NON TER 1 1
FT CHAIN <1 26 glucoamylase.
FT CHAIN 27 85 glucoamylase.
FT CHAIN 86 411 glucoamylase.
SQ SEQUENCE 579 AA; 62151 MW; 75811D6936BB8751 CRC64;

Query Match 26.3%; Score 728; DB 3; Length 579;
Best Local Similarity 37.1%; Pred. No. 2.8e-38;
Matches 166; Conservative 70; Mismatches 170; Indels 42; Gaps 10;

QY 25 ATLDWSLNSNATVARTAILNIGADGAWSGAGSIIVASPTDNPDTFTWTRDGLVL 84
Db 143 STSSWIDGQTSVRVAMLRNINP-----PGSATGFLAASLSTAGPDYFYAMTRDSALTS 197
QY 85 KTIYDLFRN---NGDTSLSSTIENYISAQAIYQGISNPSGDLSSGAGIGEPFNVDATYT 141
Db 198 NVLYVENNTLLSGNKITLAVLKDYVTFVSQTSVTCN-----CLGEPEFNKSSGYS 251
QY 142 GSWGRPQRDGPALRATAMIGFGQ--WLDNGYTSTATDIWPLVHNDSLVAQYNNQTCYD 200
Db 252 GAWGRPQNDGPASADPFIILADSLIKQTDGATYVTGTLAPAIKLDYVSTSNCGCLD 311
QY 201 LMEEVNGSSFFTLAVQHRALVEGSAFATAVGSS--CSWCDSQAPELICYLQSFN--TGSF 256
Db 312 LMEEVNGVHFYTLVWVRGVLKVGANFATRNDSFTRATYTTNTAASIKTKIDSFNNSNGQY 401
QY 259 ANFDS-----SSRGKQANTL---LGSIHFPDEACDDSTPQPCSPRALANHKEVVD 308
Db 372 IQVSQSVTGVSXKAGYASVLIAASNLGSL-----DGFPTPSEKILATAVAVEDS 422
QY 309 FRSIYTLNDGLDSEEAVALGRYPEDTY-----YNGNPFILCTLAAREQLYDALYQMDKOG 363
Db 423 FASIVPINKKVLPSYLGNSIGRYPEDTYNGNGNSQGNPFILAVTGALRYAALKEWFRNG 482
QY 364 SLEVTIVSLDFKALYSDATGT--TYSSSSSTYSIIVDAVKTFFADGFSIVETHAASNGSM 422
Db 483 GVTYTSISLNFPPFKFSSATSGKKYVGTSDFNKLONIALADRFILSTVOLHAHNNGSL 542
QY 423 SEQYKSDGDEQLSARDLTWGAALLTAN 450
Db 543 ABEFDRITTGASTGARDLTWASHASLITAS 570

RESULT 10
Q8TFE5
ID 08TFE5 PRELIMINARY; PRT; 515 AA.
AC 08TFE5;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Glucoamylase precursor (EC 3.2.1.3).
GN GUU 0111.
OS Saccharomyces fibuligera (yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
```

OC	Saccharomycetales; Saccharomycopsidaceae; Saccharomycopsis.
OX	NGBI_TaxID=4944;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=IFO 0111.
RA	Hoshino E., Soloviova A., Janacek S., Gaapetix J.;
RT	"Raw starch degrading glucoamylase from Saccharomycopsis fibuligera:
RL	molecular cloning and expression in yeast."
DR	Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
DR	EMBL: AJ11587; CAC83969.1; "
DR	GO: GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.
DR	GO: GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR	GO: GO:0005976; P:polysaccharide metabolism; IEA.
DR	InterPro: IPR001065; Glyco_hydro.15.
DR	InterPro: IPR008928; Glyco_trans.6p.
DR	Pfam: PF00723; Glyco_hydro.15; 1.
DR	PRINTS: PR00736; GLHYDRLASE15.
DR	PROSITE: PS00820; GLUCOAMYLASE. 1.
KM	Signal; Hydrolase; Glucosidase.
FT	SIGNAL 1 26
PT	CHAIN 27 515
SO	SEQUENCE 515 AA; 57423 MW; A2F27D9EBA84ZAEO CRC64;
<hr/>	
Query Match	24.3%; Score 674; DB 3; Length 515;
Best Local Similarity	33.7%; Pred. No. 6.7e-35;
Matches 163; Conservative 82; Mismatches 177; Indels 62; Gaps 16;	
OY	24 RATUDSLNSNEATVARTALINNGADGAWVGASDGIVVASPSTNPDPFYTWTRDSGLV 83
Db	41 RSNNTQWIIHEPAVSWYLLQNIIDYPEGOFKSAKGVAVASPSTSEPFYTQWTRDAIT 100
OY	84 LKTIVDLFRN--GDTSILSTIENYISAQAIVOGISNBSGDLSS--GAGLEPKENVDET 138
Db	101 FLSLIAEVEDHSFSNITLAKVEYYISNTYTLQRVSNSPNDSNDHDLGSPKENVDIT 160
OY	139 AYTSNGRPPORDGAPLRATATIGF-----GMQL--DNGYT-STATDIWPLVRND 186
Db	161 AYTSMGWGPONDGPALRAYALSRYLNAAKNNGTLLLAGONGIPYSASADIWKIKIPD 220
OY	187 LSVAAYQNVNONGYDLMEVNSSPFTIAVOHRALVEGSAFATVAGSS--CSWCSDGAKEI 244
Db	221 LQHVSTHWKSTSGFDMEBNQGTHFFITALVOLKAUSYG.PLSLTYNDDPGFTSLERQKDAL 280
OY	245 LCYLQSFWTGSFTILANPD-----SSRSCKDANTLLGSIHTPEACDSTFPQSCP 296
Db	281 NSYINS---SDFVNSCKKHVIYESPOLSRGGLDSATYAIALITHD---IGDDTYTPEN- 333
OY	297 RALANHEVDSF-----RSITYLANDGLSDSEAVAVGRYPEDTY-----YNGNPWFLC 344
Db	334 ---VDNSGYVNLISLYLLLVDNKKRYKINGNY--KAGAAGRVPEDYNGVTSEGPMOLA 388
OY	345 TLAAAEQULYDALYQMCD-QGSLETVDVSLDFPKALYSMA--TGTVSS-----SSST 393
Db	389 TAYVGGTFPYTLAVYSLKKRKVLVEKLKNLDYNSFIADLSKIDSSYASKDSLITLTYGGDN 448
OY	394 YSISIVDAKVPADGFVSIVETFAASNGSMSQCYDSDGEQSARDLTWSYALLTANNRR 453
Db	449 YKVNIKSLQCGDSFKVLKLHDHIDNGQLTEIRNKTYFOAGAVSLTMSGSSLSANRAR 508
OY	454 NSVV 457
Db	509 NKLI 512
<hr/>	
RESULT 11	
ID Q96Z10	PRELIMINARY; PRT; 599 AA.
AC Q96Z10;	
DT 01-DEC-2001 (TrEMBLrel. 19, Created)	
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)	
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE Hypothetical protein ST2017.	
NN	

OC Sulfolobus tokodaii.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
CC Sulfolobus.  
XX NCBI\_TaxID=111955;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JCM 10545 / ?;  
RX MEDLINE=21456156; PubMed=11572479;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,  
RA Nagai Y., Nishiihama K., Otsuka R., Nakazawa H., Takamiya M., Ogo Y.,  
RA Yoshitawa T., Tanaka T., Kudo Y., Yamazaki J., Kusihida N., Oguchi A.,  
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
RA Oshima T., Kikuchi H.;  
RT "Complete genome sequence of an aerobic thermocacidophilic  
RT Crenarchaeon, Sulfolobus tokodaii strain7.";  
RL DNA Res. 8:123-140(2001).  
EMBL: AP000988; BAB67116.1; "-".  
DR GO: GO:0016020; C:membrane; IEA.  
DR GO: GO:0004339; F:Glucan 1,4-alpha-glucosidase activity; IEA.  
DR GO: GO:0004872; F:Receptor activity; IEA.  
DR GO: GO:0005215; F:transporter activity; IEA.  
DR GO: GO:0005976; F:polyaccharide metabolism; IEA.  
DR GO: GO:0006810; P:transport; IEA.  
DR InterPro: IPR001615; Glyco\_hydro\_15.  
DR InterPro: IPR008928; Glyco\_trans\_6np.  
DR InterPro: IPR000531; TonB\_boxC.  
DR Pfam: PF00723; Glyco\_hydro\_15; 1.  
DR PROSITE, PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
QO Hypothetical protein; Complete proteome.  
QO SEQUENCE 599 AA; 69661 MW; F8B94CAE7731D119 CRC64;

Query Match	6.7%;	Score 187;	DB 17;	Length 599;
Best Local Similarity	22.3%;	Pred. No. 0.001;		
Matches 102;	Conservative 58;	Mismatches 151;	Indels 146;	Gaps 23;
QY	29	SWL-----SNEATVARTAIL-----	NNIGADGAMVSGADSGIVAVSPSTNDPFFYTWTR	78
		:::	:::	:::
Db	235	AMLVKARDYSIVRKRSLIIIAAHMKN--	GALPALDIDIRFNDT-----YNYVMHR	286
QY	79	DSGLVLTLDVLFPRNGDSTLLSTIENTYISAQAIYQGISNPSGDISGAGLGEPRFNVDET		138
		:::	:::	:::
Db	287	DAAF-----	ASIALTLGYG--QDEIRLNFETKPLIFNGFLFO--KYTCD--	327
QY	139	AYTGSWGRP-----	QRDGRPALRATAMIGFGQWMLDNGYTSNTATYIWPVLRNDL	187
		:::	:::	:::
Db	328	---GNMSSTHMPMNPRSIPIQED----	ETALLALYALWVHFSRPTD--IDFVAPLVAPEV	377
QY	188	SVYAQWY-----NOTG-----	YDLMEEVNGSSFFTLAVCHRALVEGSAFATAVGSSCSWC	237
		:::	:::	:::
Db	378	KLIAEFLVSYRDEDTGILPLPSYDLMERLGTHTFTSLAVYAGLSAKKAFEPFODE--NLK		436
QY	238	D---SQAREILCYLQSFMTGS--FLIANFDSSRSQK--DANTLLGSIHTFPDEAACDDSTF		291
		:::	:::	:::
Db	437	DKYLTIANEVKKGLERFYVDGDFHARTIYEDNSIDKTVADASTLPAASIL-----	GPF	486
QY	292	QPCSPRALAHKEVYDSFRSITYTLNGLSSSEAVAVGRYEDPY---	NGNPFELCTLA	347
		:::	:::	:::
Db	487	DPKQPRVISNRKVEEKL-----NINGG-----	IARYENDWYLKODEKSNMFIITLLW	534
QY	348	AAEQLYDALYQMPDKQGSLEVTDVSLDPEFKALYSDAATGTVSSSSSTVSSIVDAVKTPEADG		407
		:::	:::	:::
Db	535	LAQY-----	-----ILBGNKEAKK	550
QY	408	FVSIVETHAASNGSMSEQYDKSDGEOISARDLTWSYVA	444	
		:::	:::	:::
Db	551	YIDWIVISHMLPTGIPEQVSPKN--TYPSPVAPLWMSHA	586	
RESULT 12				
Q9HL88				
ID				
Q9HL88				
Q9HL88;				
PRELIMINARY;				
PRT;				
636				
AA.				

DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Hypothetical protein Ta0342.  
GN Ta0342.  
OS Thermoplasma acidophilum.  
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;  
OC Thermoplasmataceae; Thermoplasma.  
OX NCBI\_TaxID=2303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 1728;  
RX MEDLINE=20479972; PubMed=11029001;  
RA Ruepp A., Gaml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
Mewes H.-W., Fishman D., Stocker S., Lupas A.N., Baumeister W.;  
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma  
acidophilum."  
RL Nature 407:508-513(2000).  
DR EMBL: AL445064; CAC11486.1; -  
DR GO: GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.  
DR GO: GO:0005976; P:polysaccharide metabolism; IEA.  
DR InterPro: IPR000165; Glyco\_hydro\_15.  
DR InterPro: IPR008928; Glyco\_trans\_6hp.  
DR Pfam: PF00723; Glyco\_hydro\_15; 1.  
KM Hypothetical protein; Complete proteome.  
SQ SEQUENCE 636 AA; 73269 MW; FFA4656CCDE9AFD CRC64;

Query Match 6.7%; Score 185.5; DB 17; Length 636;  
Best Local Similarity 19.0%; Pred. No. 0.0014;  
Matches 86; Conservative 56; Mismatches 139; Indels 171; Gaps 15;

QY 57 DSGIVVSPSTD-----NPDPYFTWTRDSGLVLTVDLFRNGDPT--LLSTIENYISAO 109  
DB 281 DLGAIASCSDDLKLSHDGYYVWPADASMAAYAL-SIGHSPTARFALMDSISRE 339  
QY 110 AIYQGISNPSGDSGSGGLGEPKKNVDETATYSGWGRPODGRPL-----RATMIGFGQ 164  
DB 340 GYLYHKYNDGKIAS-----SMLPWNKSIYPIQDEDTALVYVVAL 381  
QY 165 WLDNGYSTATDVLWPLVRNDLSVYAQYNN-----OTGYDLME 204  
DB 382 WEYPRK-----NDIGFTAPYERLITRAADFMWTVNDNGPKPSFDLME 428  
QY 205 VNGSSPFTIAVQHRALVEGSAFATAVGSSGWCDSQAPILCYLQSFMTGSFTLANPSS 264  
DB 429 RYGHIAVTATVVAALKAASNFANVFC-----DPD 458  
QY 265 RSGKDATLIGSITPPPEAACDSTFQCSPRALANKE--VDSFRSITYTLNDGLSDS 322  
DB 459 LSEKYENAAERMYAFDERFYSEDTGY--ARAIDGKDPFTVDSALTSVLV-FGPKDA 514  
QY 323 E-----AVAVGRYPEDTYN-----GNPFLCTLAAEQLYDA 355  
DB 515 DDERKIVISTMRISDLWVNGVGARQYQNDRIKRVKDDPSVPGNPMITITLMAH-RYYMR 573  
QY 356 LYQWMDKQSLVETVDSLDFPKALYSDAATGTYSSSSSTYSIVDAVATFADGFVSIYETH 415  
DB 574 FGPPEKAMNL-----IQWKS 590

RESULT 13  
ID 08XD4 PRELIMINARY; PRT; 1588 AA.  
AC 08XD4;  
DT 01-MAR-2002 (Tremblrel. 20, Created)  
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Putative adhesin (Hypothetical protein).  
GN Z5029 OR ECS4480 OR H161.

OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
Rasofai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohemorrhagic Escherichia coli O157:H7";  
RL Nature 409:529-533(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Toke T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kihara S., Shiba T., Hatori M., Shingawa H.;  
RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:11-22(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RIMD 0509952;  
RA Makino K., Yutsudo C.H., Yokoyama K., Kubota Y., Kimura S.,  
RA Shingawa H.;  
RT "O157 specific gene similar to H. influenzae adhesin gene";  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB005587; AAG58749.1; -  
DR EMBL: AB002566; BAB37803.1; -  
DR EMBL: AB036416; BAB87814.1; -  
DR PIR: A86036; A86036.  
DR PIR: H91188; H91188.  
DR InterPro: IPR008640; Hep\_Hag.  
DR InterPro: IPR008635; HIM\_  
DR InterPro: IPR005594; Yada,  
DR Pfam: PF05658; Hep\_Hag; 13.  
DR Pfam: PF05662; HIM; 12.  
DR Pfam: PF03895; Yada; 1.  
KM Hypothetical protein; Complete proteome.  
SQ SEQUENCE 1588 AA; 160150 MW; B2BA4E06F28DEC CRC64;

Query Match 6.5%; Score 181; DB 16; Length 1588;  
Best Local Similarity 21.7%; Pred. No. 0.0092;  
Matches 136; Conservative 95; Mismatches 217; Indels 178; Gaps 30;

QY 15 TGLANYSKRALDLSNLSEATATVAFALINNICADG-AWMSGADSGIVVSPSTNDPDR 73  
DB 576 TTNANNTSNIAITWTNISN-----LIEVTNLGEDALKM--DKONGFTNAGTETTSKI 628  
QY 74 YTWTRD-----SGVLKTLVDLFRNGDPTSLSTIENYISAQAIYQGISNPSGDL 122  
DB 629 -TNVXGDLITGSGTDVANGSLKTTNDAYANTTNIAINTN-----ISNLTEIV 677  
QY 123 SSGAGLGEF--KENVDETATYSGWGRPQRDGPALRATAMIGQGMILLNDGYSTATDIW 180  
DB 678 TN---LGEDALKVDKNDNGVFTAHG-----NNTASKITN-----LIDGTVATSSDAIN 723  
QY 181 PLVRNDL-SVAYQW-----NOTG-----YLMEEVNGSSFTIAVQHRALVEGSAFA 227  
DB 724 GSQLYDLSSNIATYFGGNASVNTDGVFTGFTYKIG-----TNYNVGDALAAI--NSSFS 777  
QY 228 TAVSSCSWCDSQAPILCYLQSFMTGSFI--LANPDSRSRGDA--NTLLGSIHTF--- 280  
DB 778 TSLGDALW-DATAGKFSAGKGTNGDASVITVDADDEISDSSDANVNGSLDHGVSSVVD 836  
QY 281 ---DPEAACDSTFQCSPRALANKEVVDSFRSI-YTLNDGLSDSEAVAVGRYPEDTY 335

Db 837 ALGGAENVADGTTTAPTYTIANADYDNDALNADITLDDAL-----LMDADAG 887  
QY 336 YNGNPFLCTLAABEQLYDALYQMDKQSLFVYDVSLDFKALYSDAAT-----385  
Db 888 ENG-----AFSAHKGDKTASV-ITNVANGAISAAASDAINGSOLYTTNKYI 933  
QY 386 -----TSSSSSTYSSTIVDAVKT-----A 405  
Db 934 ADALGGDAENVADGTTTAPTYTIANAEYNNVGDALDLDNALMDETANGAGAYNASH 993  
QY 406 DGFVSIETHAASNGSMSE-QYDKSDGEQLA-----RDLTWSYAL 446  
Db 994 DGVASIIITN--VANGSISESDTDVANGSQANATMMIEQNTQIINQLAGNTDATYIQENG 1051  
QY 447 LTNRRNSVVPASMETASSTSPGTCATSAICTYSSVTWSPSIVATGCTTTATPT 506  
Db 1052 AGINIVRTDGDGLAFDASQGVGAT-----AIG-YNSVAKGDSVAIGGYSVDVTGI 1105  
QY 507 GSGSVTSTKTTATASKTSTTTSGM 532  
Db 1106 ALGSSSVSRVIAKGSRDISTENG 1131

RESULT 14  
Q9C105 PRELIMINARY; PRT; 1236 AA.  
AC 09C105;  
DT 01-JUN-2001 (TEMBLrel. 17, Created)  
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
DE Putative glucosylase I (Alpha-1,4-glucan glucosidase), extracellular starch-degrading enzyme, by similarity to S. cerevisiae STAl, contains chitinase family signature.  
GN SPAP1E7.04C.  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
NCBI\_TaxID=4896;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=972h-;  
RA Wood V., Rajandream M.A., Barrell B.G., Seeger K., Harris D.;  
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL/AL590605; CAC36921.1; -;  
DR GeneDB\_Spombe; SPAP1E7.04C; -;  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0008152; P:metabolism; IEA.  
DR InterPro; IPR001917; AminoTrans-II.  
DR InterPro; IPR001223; Glyco\_hydro\_18.  
DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
DR PROSITE; PS00599; AA\_TRANSFER\_CLASS\_2; 1.  
SQ SEQUENCE 1236 AA; 123387 MM; 5A2D3JA30B87CDD8 CRC64;

Query Match 6.3%; Score 175; DB 3; Length 1236;  
Best Local Similarity 21.5%; Pred. No. 0.016; Matches 199; Conservative 91; Mismatches 242; Indels 138; Gaps 21;

QY 10 SGVLCGLANVISKRALDLSWLSNEATVARTALINNIGAD-----49  
Db 528 SSISSPMNSISSSSSIPTSSSDPSSTITSSGISSSISPTSSSVSISSSSSP 587  
QY 50 ---GAVSGADSGIVVASTNDPDPFYFTRDGLVLTVDLFRNGDTSLSSTIENYI 106  
Db 588 SSTSLSSST 637  
QY 107 SAAATVQGSN-----PSGD--LSSGAGLGEPKRVNDETATY 141  
Db 638 SSSSISSSSSISSSISSSIPTSSSLSTYSSSVIPSSSTVSSSSSL-----IVSSSPA 692  
QY 142 GSWGRFQRDGPALRATAMIGFQWLLDNGYSTATDVIWPLVANDLSVYAQVNNQCYDL 201

Db 693 SSSSSPISPASSISVST-----YSASLSNI-----THSSLSLRAM-----SSSSAI 732  
QY 202 WEVNSSSPFTIAVCHRALVEGSAFATVAGSS--GCWCSQAPFELICYSFPTGSPITIA 259  
Db 733 PTVNSSTLIT-ASSNTLLSSITSSSAIVSSSTVSNISNIPSAVASSQSOLNSSTLA 791  
QY 260 N--FDSSRGKANTLLGSIHFPDPAACDDSTFOCSPRALNKHVNDSPRSITLND 317  
Db 792 TSLYLSSSSSRITST 848  
QY 318 GLSDSAVAVGRYPEDTYNNGNPFCLTLAAEQLYDALYQMDKQSLFVYDVSLDFKA 377  
Db 849 -LESTSVSTATYTDSVTS-----TTLATSQ-----GSSSVSSSLSTTS 890  
QY 378 L-----YSDAAGTYSSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 414  
Db 891 LSTSIPTSSVAPAVTST 950  
QY 415 HAASNGSMEQYDKSDGEQLSARDLTWSYAALLTNRRNSVVPASMGTSASSVPG---471  
Db 951 SSSSSSSSEVDPTAASATGSS-TSSITATASVSGSSSTSVATASATDSSTSIASA 1009  
QY 472 TCAATSAICTYSSSVTVSWPSIVATGCTTTATPTSGSVT-STSKTTATASKTSTTRS 530  
Db 1010 TGSSTSVAT-ASVTDSTSTSVATASATDSSTSIASAATGSSSTSVATASATDSST 1068

RESULT 15  
Q97BM7 PRELIMINARY; PRT; 659 AA.  
AC 097BM7;  
DT 01-OCT-2001 (TEMBLrel. 18, Created)  
DT 01-OCT-2001 (TEMBLrel. 18, Last sequence update)  
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
DE Glucoamylase.  
GN TV0428 OR TVG0415361.  
OS Thermoplasma volcanium.  
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;  
OC Thermoplasmataceae; Thermoplasma.  
NCBI\_TaxID=50339;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=GSS1 / DSM 4299 / JCM 9571;  
RX MEDLINE=20570466; PubMed=11121031;  
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,  
RA Kawashima T., Yamamoto Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,  
RA Nunoshida T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;  
RT "Archaeal adaptation to higher temperatures revealed by genomic  
RT sequence of Thermoplasma volcanium."  
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262 (2000).  
DR EMBL; AF000992; BAB59570.1; -;  
DR GO; GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.  
DR GO; GO:0005976; P:polysaccharide metabolism; IEA.  
DR InterPro; IPR00165; Glyco\_hydro\_15.  
DR InterPro; IPR008928; Glyco\_trans\_6hp.  
DR Pfam; PF00723; Glyco\_hydro\_15; 1.  
KM Complete proteome.  
SQ SEQUENCE 659 AA; 76598 MM; 36101F5405F13B33 CRC64;

Query Match 6.3%; Score 174.5; DB 17; Length 659;  
Best Local Similarity 19.7%; Pred. No. 0.0071; Matches 90; Conservative 49; Mismatches 158; Indels 159; Gaps 16;

QY 57 DSGIVVASTD-----NDPYFYTWRDGLVLTVDLFRNGDTSLSSTIENYISAQAI 111  
Db 304 DLGAIAASCDSDILKSHDGYVWPDPASIAAYAL-----SIGSHSETPARFPAL 354  
QY 112 VQGISNPSGDLSSGAGLGEPKRVNDETATYGSV-----GRP-----QRDGPALRATAMIG 161  
Db 355 MESSMSEEGYLH-----KNVD-GKIASSLPHVINGKHVYPIQEDETAL-----399  
QY 162 FQWLLDNGYSTATDVIWPLVNR-----NDLSVYAQVNN-----Q 196

[illegible]

Query Match	6.3%	Score 174.5	DB 16	Length 990
Best Local Similarity	21.3%	Pred. NO.0.012		
Matches 128	Conservative 93	Mismatches 201	Indels 179	Gaps 28
QY	43	LNNIGADG-AMVSGADSGIVVASEPTNDPFIYTWTRD-----SGLVLEKTLVLD	90	
DB	1	MTNIGEDALKW--DKDNGVFTPAHGTETTSKI-TIVKXGDLLTGTDAVNSQLTWTDA	57	
QY	91	FRNGDTSLSLTENYISQAATVQGISNPSGDLSSGAGLEP--KENVDETAYTGSWGRPQ	148	
DB	58	VAINTNTTATNTN-----TSNLTETVTN--LGEALKWDXKNGVFAHG--	101	
QY	149	RDGPALRATAMIGFGQMLDNGYSTATDVIPLVRNDL-SYVAQYV-----NOTG---	198	
DB	102	NNTYASKITN-----ILDGIVYATTSDAINGSLVYLSNIAIATYFGNANVATVDGFT	153	
QY	199	---YDLWEVNGSSFTTIAVOGHRALVEGSAFATVAGSSCSWCDSCAPEILCYLOSFWTGS	255	

```

Db      154 GPYYKIG-----TNYVVGDALAAI--NSSFSTSLGDALLW-DATAGKRSAGHGTMGDAS 206
Qy      256 FI--LANPDSSRSRSGKA--NTLLGSIHTF-----DPEAACDDSTFOCPSPALLANHE 304
Db      207 VITPDADGEISDSSDAVNGSQLHGVSVVVDALGGAENVNDGTTTAPTYTIANADYUN 266
Qy      305 VVDSRSLTYTLANDG--LSDSEAVAVGRPEDETYNGNPFELCTIAAEQYDALLYQWDCQ 362
Db      267 VGDALNALADITPPDDALLMDADAGENG-----AFSAAHKQDKT 303
Qy      363 GSELEVTVDLSDFFKALYSDDATG-----TYSSSSSST 393
Db      304 ASV-ITTVANGAISASSDAILNGSLYTNNKXIADALGGDAEVNADGTTTAPTYTIANNE 362
Qy      394 YSISIVDAKTF-----ADGFVSIETHAASNGSMSE-QYIKSDG 431
Db      363 YNNVGDDALDALDNNALLMDKTANGAGAYNASHDGEASIIITN--VANGSISIEDSTDAVNG 420
Qy      432 EOLSA-----RDLTWSYALLTLTANNRNSVVPRAWSSETSASVPGTC 473
Db      421 SOLNNTNMMIEONTOTIINOLAGNTDATTYIEENGAGINIVYRTDNDLIAFDADASGVGATA 480
Qy      474 AATSALIGTYSSTVYVSWPISIVATGTTTATPFG--SGSVTISKTTATATASKSTYTRSG 531
Db      481 VGVNA-----VASGASSVAIGNSSSTVDGTGIALGSSSVSSRYTLAGSRDTSVTEMG 532
Qy      532 M 532
Db      533 V 533

```

ID	Q8TGE1	PRELIMINARY;	PRF;	1713 AA.
AC	Q8TGE1;			
DT	01-JUN-2002 (TRENBLREL, 21, Created)			
DT	01-JUN-2002 (TRENBLREL, 21, Last sequence update)			
DT	01-JUN-2003 (TRENBLREL, 24, Last annotation update)			
GN	Cell wall protein Awalp.			
GN	AWAL.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=2191416; PubMed=11916725;			
RA	Shimoi H., Sakamoto K., Okuda M., Atchi R., Iwashita K., Ito K.;			
RT	"The AWAL gene is required for the foam-forming phenotype and cell			
RT	surface hydrophobicity of sake yeast.";			
RL	Appl. Environ. Microbiol. 68:2018-2025(2002).			
DR	EMBL; AB071164; BAB85832.1; -			
DR	GO; GO:0006950; P:response to stress; IEA.			
DR	InterPro; IPR000992; SRP1.T1P1.			
DR	Pfam; PF00660; SRP1.T1P1.1.			
DR	PROSITE; PS00724; SRP1.T1P1.1.			
SQ	SEQUENCE 1713 AA; 16883 MW; 9321A296B586404 CRC64;			
Query Match	6.3%; Score 173.5; DB 3; Length 1713;			
Best Local Similarity	22.6%; Pred. No. 0.031;			
Matches 125; Conservative	91; Mismatches 281; Indels 55; Gaps 15;			
QY	7 LALSGLVCTGLANVTSKRAITDLSLN-EATVATATALLNIGADGAVSGADSGIVASP 65			
DB	393 LAVSTVSTSTSDSVSPNIPSEISSPESSTALTATSSGSSKATSGSSVSGTSTSGSS 452			
QY	66 STDNDPYEYTWTRSDGLVTLKTVLDFRNGDTSL-----SPTEYVISAQAIQISNPSG 120			
DB	453 ASGSS-----ATSGSSVSGSTGATSGASASSVSSGSGVTEGSSSSASGSSITQ-SG 506			
QY	121 DLSSGAGIGEPKENVDE-TAYTSKGR-----PQRDPLKATPAIIGFQWLND 169			
DB	507 -TAGGSSVTSSTGSGVLTGSGSSASASAPGISSTIPGTSSTASSTAGSTSGTSTGTSIT 565			



Dd QY 170 GYSSTTDLVWFLVRNDLSLVQAYNNNGDYDLMEEVGSSPFTJAVOHRALVEESAPATA 229  
|||::||  
Db SGGSSATTE-----SGSVGSSSSTATEBGSSVS---TSATEGSVSGSTATE 612  
|||:  
QY 230 VGSSCWCMDQAPELLCYLQSFWTGSLFANFDSSRSRGRANDTLGLIHFPDEACDDG 289  
|||:  
Db SGSSAGSSSATNES-----GSSVCGSTSTATBGSVS GSTATBGSAGSSSATNBGS 667  
|||:  
QY 290 TTP-PCSPALANKHEVDPSFRISITLTLDGLSDSAVAVGKYPEDITYNGNPFCLTAA 348  
|||:  
Db ASSVPSSSSSVTESGSSSSASBSSTIGCTA GSASSATS--GSVTQGSGSVGSSASS 724  
|||:  
QY 349 AEGLYDALQWDKQCGLSETVTDVLDFPALXADATGYSSSSSYSSIYDAVTFADF 408  
|||:  
Db APGISASTIQSTSMASTBASGTTCGLTITSISSATESGSSASGSSSATBESSVS 784  
|||:  
QY 409 VSIVETHAANSMSBOYDKSDGEOLASRDILWSYAALLTNRRNSVPAAMGET---S 465  
|||:  
Db TSATBEGSSVSGSTATBEGSSASGSSSATBEGSSVSGSTATBEGSSAGSSSATBEGS 844  
|||:  
QY 466 ASNVPGTCAATAIIGYSSVYTSMNPVIATOGTTTLPPIGSGSVTSKTYTAASKTS 525  
|||:  
Db ASSVPSSSSSVTESGSSSS---ASSESTITQGTAGSSASBSTSGSVTOQSGSVGSSASS 901  
|||:  
QY 526 TT---TRSGMUL 534  
|||:  
Db TGSGVTOGSGSV 913

RESULT 18	
Q97ZD0	
ID Q97ZD0	PRELIMINARY;
	PRT; 622 AA

DT 01-OCT-2001 (Tremblurel, 18, Created)  
DT 01-OCT-2001 (Tremblurel, 18, Last sequence update)  
DT 01-OCT-2003 (Tremblurel, 25, Last annotation update)  
DE Glucan 1,4 alpha glucosidase (Glucosylase) (EC 3.2.1.3).  
GN SSO0990.  
OS Sulfolobus solfataricus.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
OX NCBI\_TaxId=2287;  
[1]  
RP SEQUENCE FROM N.A.  
RC SRRAIN-ATCC 35092 / DSM 1617 / P2;  
RX MEDLINE=21332286; PubMed=11127726;  
RA She Q., Singh R.K., Contoliveri F., Zivanovic Y., Allard G.,  
RA Aayaz M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,  
RA De Moors A.J., Erauso G., Fletcher C., Gordon P.M.K.,  
RA Heikamp-de Jong I., Jeffries A.C., Kozeira N., Peng X.,  
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstunov N.,  
RA Charrelot R.L., Doolittle M.F., Duguet M., Gaasterland T.,  
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.,  
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.",  
RT Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
RL EMBL: AE006718; AAK41262.1; -  
RL PIR: G90250; G90250.  
DR CO: GO:0004339; F:Glucan 1,4-alpha-glucosidase activity; IEA.  
DR CO: GO:0016798; F:hydrolyase activity, acting on glycosyl bonds; IEA.  
DR GO: GO:0005976; P:polysaccharide metabolism; IEA.  
DR InterPro: IPR000165; Glyco\_hydro\_15.  
DR InterPro: IPR008928; Glyco\_trans\_6tp.  
DR Pfam: PF00723; Glyco\_hydro\_15; 1.  
DR Hydrolase; glycosidase; Complete proteome.  
KW SEQUENCE 622 AA; 72052 MW; C7B411FD231AB8F CRC64;  
SQ

Query Match	6.2%	Score 172	DB 17	Length 622
Beet Local Similarity	21.7%	Pred. No. 0.0095		
Matches 97	Conservative 56	Mismatches 165	Indels 130	Gaps 19
QY	37	VARTAIINNICADGAWSGADSGIVVAAPESTNDPQYFTWTRDSGLVKTLYDLEFRNGDT	96	
	:   :   :   :   :			:

```

Db 268 VSLFVIRNHMDVNSIIASSDFFSV----K1YGD9YQYCPWDAIAAVAL-DLNGYKEL 322
Qy 97 SL--LSTTENTYSAOAVIQGISNPSGDLSGAGLGEKPFVNDETAYTGSWGR-ORDGPA 153
Db 323 AKKHQFQISNANSEGFLYKHKNPPTLASS---WHPVY-----YKGRIRYIQEDETA 373
Qy 154 LRATAMIGFGWLMDNGYTSTATDI-----VWLVVRNDLSYVAQYNNQGYDIME 203
Db 374 LEV-----MAIASHY-EKYEDIDEILPLYKKFVKPALKPMFMSFMEGGLPKPSFDLWE 424
Qy 204 EVNASSPFTIAVQIRALVEGSAFATAVSSSCGSCDSQAPBILCYLQSFMTGSFILANPFS 263
Db 425 ERYGIHIHYSTVYGALTKGAKLAYVDGE-----LLESDIS 461
Qy 264 SRGSKDANTLLGS-----IHTDPR-----AACDSTFQPCSPRAL-ANHKREVDSFRS 311
Db 462 DTSGILKGMVLRKMTYNGRPFVRRIDEENNODLTVDSSSLYAPFPGLVANDKIMITINE 521
Qy 312 I---YTINGLSDSEAVAVGRYPRDYTY---NGNPWFICTLAAAEQLDALYQMKQGS 364
Db 522 IESRLTVNGGII-----RYENDMYORRKKQPPMWTITLV-----W----- 566
Qy 365 LEVTDVSLDFPKALYSDAATGYSSSSSTGYSIYAVTPEADGFSIVETHAASNGSMSE 424
Db 557 -----LSEYYATIND-KKKANEYIKWVIRRALPTGFLPE 589
Qy 425 QYDKSDGEQLSARDLTWMSYALLTFANNR 452
Db 590 QVDETEPEPFSVTPLWMSHAEPITAIANK 617

```

RESULT 19  
Q88XB6  
ID Q88XB6 PRELIMINARY; PRT; 3360 AA

DT 01-JUN-2003 (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Cell surface SD repeat protein precursor.  
GN SDR OR LP 1303.1 OR LP 1303A.  
OS Lactobacillus plantarum.  
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
OC Lactobacillus.  
OX NCBI  
OX taxid=1590;  
RN [1]  
RN SEQUENCE FROM N.A.:  
RC STRAIN=NCIMB 8826 / WCFS1;  
RX MEDLINE=22480296; Pubmed=12566566;  
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,  
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,  
RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,  
RA Hoffer S.M., Nispor Groot M.N., Kerkhoven R., De Vries M., Ursing B.,  
RA De Vos W.M., Siezen R.J.;  
RT "Complete genome sequence of Lactobacillus plantarum WCFS1.",  
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995 (2003).  
DR EMBL; AL935255; CAD63798.1; -  
DR InterPro; IPR003343; Big\_2.  
DR Pfam; PF02368; Big\_2; 1.  
KW Complete proteome.  
SQ SEQUENCE 3360 AA; 332641 MW; BBF2411DF439F446 CRC64;

[illegible]



RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPX101-1C;  
 RX MEDLINE=89031230; PubMed=3141213;  
 RA Paro J.M., Ianez E., Zalacain M., Clares M.G., Jimenez A.;  
 RT "Similar short elements in the 5' regions of the STR2 and SGA genes  
 from *Saccharomyces cerevisiae*.";  
 RL FEBS Lett. 239:179-184(1988).  
 DR EMBL; X13858; CAA32071.1;-.  
 DR HSRP; P08017; IANY.  
 DR GO; GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.  
 DR GO; GO:0005976; P:polysaccharide metabolism; IEA.  
 DR InterPro; IPR000165; Glyco\_hydro\_15.  
 DR InterPro; IPR008928; Glyco\_trans\_6np.  
 DR Pfam; PF00723; Glyco\_hydro\_15; 1.  
 DR PRINTS; PR00736; GLHTRLASE15.  
 FT NON\_TER 190 190  
 SQ SEQUENCE 190 AA; 20902 MW; 3799938DB287F40E CRC64;  
  
 Query Match 6.1%; Score 169.5; DB 3; Length 190;  
 Best Local Similarity 30.8%; Pred. No. 0.0027;  
 Matches 44; Conservative 25; Mismatches 59; Indels 15; Gaps 5;  
  
 QY 1 MSFRSLALSGVCTGIANVTSKRATLDSWLSNEATVARTAILNIGADGAWVSGADSGI 60  
 Db 56 VQLRDVAVLNGVTVVD-SNGAMDSALAEWLGQKXVSIKIFENIGPSAVPS-ISPQV 113  
  
 QY 61 VVASPTNDPFFYTWTRDSGLVLTVDLFRNGDTSLSTENYISAQIVGGINPSG 120  
 Db 114 VIASPFQTHDPYFQWIRDSALTSINSIVS---HSAGPALETLLQYLVNSFHLQRSNN--- 167  
  
 QY 121 DLSSGAG-----LGEPRKRVND 136  
 Db 168 TLGAGIGYNDVVALGHPKRVND 190  
  
 RESULT 22  
 08XOP2 PRELIMINARY; PRT; 4106 AA.  
 AC 08XOP2;  
 DT 01-MAR-2002 (TReMBLrel. 20, Created)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Putative hemagglutinin/hemolysin-related protein.  
 GN RSP180. OR RS05070.  
 OS *Ralstonia solanacearum* (Pseudomonas solanacearum).  
 OC Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Ralstonia.  
 NC NCB1\_Taxid=305;  
 RX NCBI\_Taxid=305;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Arlet M., Billault A., Brotier P., Camus J.C., Catolico L.,  
 RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lave M., Moisan A., Robert C., Saurin W., Schlex T.,  
 RA Signer P., Thebaud P., Whalen M., Wincker P., Levy M.,  
 RA Weisenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*.";  
 RL Nature 415:497-502(2002).  
 DR EMBL; AL646083; CAD1831.1;-.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR GO; GO:0015070; F:toxin activity; IEA.  
 DR GO; GO:0003700; F:transcription factor activity; IEA.  
 DR GO; GO:0004812; F:tRNA ligase activity; IEA.  
 DR GO; GO:0006418; P:amino acid activation; IEA.  
 DR GO; GO:0009405; P:pathogenesis; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR InterPro; IPR001433; Hem|yn Ca\_bind.  
 DR InterPro; IPR000847; HTH\_LYER.  
 DR InterPro; IPR003995; RtxA.  
 DR InterPro; IPR001412; tRNA-synt\_1.  
 DR Pfam; PF00353; hemolysincbind; 4.  
 DR PRINTS; PR00313; CABDNCRPT.  
 DR PRINTS; PR01488; RTXTOXINA.  
 DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; 12.  
 DR PROSITE; PS00044; HTH\_LYER\_FAMILY; 1.  
 KW plasmid; complete proteome.  
 SQ SEQUENCE 4106 AA; 395091 MW; 6338E46766709B46 CRC64;  
  
 Query Match 6.1%; Score 168; DB 16; Length 4106;  
 Best Local Similarity 21.6%; Pred. No. 0.23;  
 Matches 137; Conservative 70; Mismatches 216; Indels 210; Gaps 29;  
  
 QY 25 ATLDSWLSNEATVARTAILNIGADGAWVSGADSGIVASPSPTNDPFFYTWTRDSGLV 84  
 Db 1679 ATAAFTLVDTAAAPAPPIIGTVTDVAPVPGA---ITAGSTNDATPTLTGAFA--- 1730  
  
 QY 85 KTLVDFRNGDTSLSTIE-----NYISAQIVGGINPSGSLSGGLSP-----K 132  
 Db 1731 NSTVSIIF--DGTTLGTTADAGSNVYPTTALTDGSHSLATATDAAGNVSPATSAFT 1788  
  
 QY 133 FNVDETA-----YTG---SWGRPQRDGPALRATA-----MIG 161  
 Db 1789 LVYDTAAAPATPVISANTDAVAPVTCITIGSGSTNDAAPTLTGAENSTTNVFDGTTLLG 1848  
  
 QY 162 -----FGQW-----LLDNGY--TSIATDIWPLVRNDLSYVAOYMNQGYDLMEVN 206  
 Db 1849 TTADALGSMWTFPTTALTDGSHSLATATD-----AAGVNS-----T 1886  
  
 QY 207 GGSFETIAVQHRLVAGSAPATVAGSSCMQSQAPETLICYLQSFMTGSPRI----- 257  
 Db 1887 ASSAFLLT-----VDIAAPATPVIGTVI--DAVAV-----TGTTAGGSTNDAN 1929  
  
 QY 258 -----LANPDSRSRGDANTLLGSI-----HTFDPPEACDSS--TFQCSPPALANHK 303  
 Db 1930 PTLTGIAEGNSTVYSIDGTTLLGTTADAGSNVYPTTALTDGSHSLATATDAAGN-- 1987  
  
 QY 304 EYVDSFRSIYTLNDGLDSEAVAVGRYPED----- 333  
 Db 1988 --VSSASAFLLTVDDAAPAAAPVIGTVDDVAPITGTVAAGSTNDTPTLTGAENST 2045  
  
 QY 334 -TYNNGNPFELCTLAABQLYDALYWDKQSLFETVDVSLDFKALYSDAATGYSSSS 392  
 Db 2046 INVFDGTTLLGTTA-----DASGNWTFPTSPPLTDG--HSPTAATADA--GNVGTASS 2097  
  
 QY 393 TYSSIVDAVYKTPADGFSIVETHAANGSMSEQYDKSDGSLARDLWLSYALLTANNR 452  
 Db 2098 AFTLVDTDAPTTPVISTVYTDVAPVPGA-----ITAGGS 2132  
  
 QY 453 RNSVPEASWGETSASVPGTCATSAIGYSSVTYTSW---PSIVATGCT---TTATP-P 505  
 Db 2133 TNDAMVLVGTAEANSTISVFDGTTLLGTTADAGSNWTFPTTALTDGSHSLATATDP 2192  
  
 QY 506 TSGGSVTS-----TSKTTATASKTSTTTSGMS 533  
 Db 2193 AGNVSTTSAFALTVDTPAAPAVITTVDAVS 2225  
  
 RESULT 23  
 09BITO PRELIMINARY; PRT; 2016 AA.  
 AC 09BITO;  
 DT 01-JUN-2001 (TReMBLrel. 17, Created)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)  
 DE Fibroin 3 (Fragment).  
 OS *Plectreurys tristis* (Spider).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;  
 OC Araneomorphae; Haplogynae; Plectreuridae; Plectreurys.  
 NC NCB1\_Taxid=33319;



RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
RA Martins E.C., Melandri J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
RA Spindola L.A.F., Takita M.A., Tamura R.B., Teixeira E.C., Tezza R.I.D.,  
RA Trindade dos Santos M., Trufi D., Tsai S.M., White F.F.,  
RA Setubal J.C., Kitajima J.P.,  
RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
RT host specificities";  
RL Nature 417:459-463(2002).  
DR EMBL: AB012148; AM39823.1; -  
DR GO: GO:009279; C:external outer membrane (sensu Gram-negative. . .; IEA.  
DR InterPro: IPR000258; Ice\_nucleatn.  
DR Pfam: PF00818; Ice\_nucleation; 68.  
DR PRINTS: PR00327; ICENUCLEATN.  
DR PROSITE: PS00314; ICE\_NUCLEATION; 40.  
KM Complete proteome.  
SQ SEQUENCE 1333 AA; 131306 MW; 326078458D0E4842 CRC64;

Query Match 5.9%; Score 163; DB 16; Length 1333;  
Best Local Similarity 23.6%; Pred. No. 0.1;  
Matches 117; Conservative 55; Mismatches 235; Indels 154; Gaps 28;

QY 38 ARTAIINIGADGAWVSGADGIVASPTDNDPYFYTWRDGLVLTLDLFRNGDTG 97  
DB 394 ARTGSDLTAGYGSTAGADSTLIAG-----YGSTQTS-----GDS 431  
QY 98 LUSTIENIYSAQIVOGISNPSGDLSSG-----AGLGEPRNVDETAYTGS 143  
DB 432 LTA--GYSTQTAARKG-----SDLTYTGSTAGADSTLIAGYGSTQTSSESSLIAG 483  
QY 144 WGRPQ--RQGPALRA-----TAMIGF-----GQWLNGYSTATDVLWPL 182  
DB 484 YGSTQTAARKGDLTLAGYGSTAGADSTLIAGYGSTQTSSESSLIAGYSTOT----- 537  
QY 183 VRNDLSYVAQYNNOT-----GYDLMEVNGSSFTIAY-QHRALVEGSAFATAVG 231  
DB 538 ARKGSDLTAGYGSTAGGSDTLVAGYGSTQTSQDSLTAGYGSTQTAARKGSDLTGTYG 597  
QY 232 SSCSWDCSQAPELCY--LQSFMTGSFILANPDS--SRGKANLTLGSIHFDEBAC 286  
DB 598 ST-STAGGSDTLTAGYGSTQTSQNASSLTAGYGSTQTAARKGSDLTGTYGSTST-AG 651  
QY 287 DDSTFQF--CSPRALANHKEVDSFRSIYTLNDGL-----SDSEAVAVAGRYPEDTYN 337  
DB 652 ABSTLIAGYGSTQTSSESSLTAGYGSTQTAARKGSDLTGTYGSTTGA-----DTLLA 706  
QY 338 GNPWFLLCTLAALYQMDKQGLEVDVSLDFKALYSDAATGYSSSSSTYSSSI 397  
DB 707 G--YGSTQTAGGESSLTAGY-----GSTQTAARKGSDLTAGYGSTTGAAGSSSLIAGGS- 758  
QY 398 VDAVATFPAQFVIV-----ETHAASNGM-----SEQYKSDGEOLSARD----- 438  
DB 759 -----TQTAGFKSLITTYGSTQTAGGSLTAGYSSSTAGDSSSLIAGYGSTQAGFK 813  
QY 439 --LTWGYAALLTANNRNRYVPAWGETSASSVPGCAATSAIGTYSSTVTSWPSIVAT 496  
DB 814 SLITAGYGSTQTAOER--SLITTYGSTTGAHDST-----LIAGYGSTQTAGKSLIT 866  
QY 497 G-GTTTATPT-----GSGSVTSTKTTATATASKTSTTTTSG 531  
DB 867 GYGSTQTAGGESSLIAGYSSSWAGPDSLIAGYGSTQTAG 907

RESULT 26  
Q9CH86 PRELIMINARY; PRT; 1063 AA.  
AC Q9CH86;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Unknown protein.  
GN YIHD OR L10852.

OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=1360;  
RN [1] \_TaxID=1360;  
RP SEQUENCE FROM N.A.  
RC STRAIN=111403;  
RX MEDLINE=21235186; PubMed=11337471;  
RA Bolotin A., Wincker P., Mauer S., Jallion O., Malarre K.,  
RA Weisenbach J., Ehrlich S.D., Sorokin A.;  
RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
RT lactis ssp. lactis 111403";  
RL Genome Res. 11:731-753(2001).  
DR EMBL: AB006319; AA004950.1; -  
DR PIR; D86731; D86731.  
KM Complete proteome.  
SQ SEQUENCE 1063 AA; 104547 MW; 5727AD9C69669A1 CRC64;

Query Match 5.8%; Score 161; DB 16; Length 1063;  
Best Local Similarity 22.3%; Pred. No. 0.1;  
Matches 132; Conservative 90; Mismatches 227; Indels 142; Gaps 24;

QY 9 LSGIVCTGLANYSKRA--TLDSMLSEATVART-----ILNITGADGAWVSGADSGI 60  
DB 425 LAGLVNNSAIVRGQIAMSQLSPTISNIPVGTNNVLSPTLND-----TGASLGE 477  
QY 61 VVASPTDNDPYFYTWRDGL---VLKTLVLDLFRNGDTSLSTIENIYSAQIVOGISN 117  
DB 478 VA-----NLTGVSLIDQVNSLGNILISGTA-LATYEN-----TLQNSLN 518  
QY 118 PSGDLSSGAG-----LGEPRNVDETAYTGSWGRPQRPQGPALRAMIGQWLMD 168  
DB 519 SFQNLPAQASDLINQVLQNALNNIVESA-TGI-----VNNLPGLG-AIE 563  
QY 169 NGTSTATDI--WPLVRNDLSYVAQYNNQGYDLMEVNGSSFTIAYQHRALVEGSA 225  
DB 564 NGLSNTISQIPNNINFNANLNGITITN---SLTPSVGAST-----VNPNSANSQ 614  
QY 226 FATAVSGSCWDCSQAPELCYLQSFMTGSFILANPDSRSGKANTLTLGSIHFDEBAC 285  
DB 615 SSSSASSSSSAAASST-----SSSNVSNSTSSNS--EANTSSSTSSASSSS 661  
QY 286 CDDSTFQF--CSPRALANHKEVDSFRSIYTLNDGLSDSEAVAVAGRYPEDTYN 343  
DB 662 SESSSASSSSSSSVASSSVDSGSSAGVASSSSSASGSSASSNSSE----- 712  
QY 344 CTLAALYQMDKQGLEVDVSLDFKALYSDAATGYSSSSSTYSSSI-----V 398  
DB 713 -SSVASSSSVDS--SQSSAGVNGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 769  
QY 399 DAVKTFADG---FVSIETHAASNGSMEOYDKSGEOLASARDLTWYAAALLTANNRN 454  
DB 770 NSSSSSABSSSASSSSSSSSSTASSSVGSSGSSSTGSSSABGSSSHS--SSNSSES 827  
QY 455 SVVPAWGETSASSVPGCAAT-----SAIGTYSV----- 485  
DB 828 SVASSSVDSGSSSAGVASSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 887  
QY 486 --TVTSWPSIVATGTTTATPTGSGSVTSTKTTATATASKTSTTTTSG 534  
DB 888 SSVSSSSSSSSSSSGTSDSNNTGQTSTSSSSKNSHSLSSNBSASSL 938

RESULT 27  
Q8VQ99 PRELIMINARY; PRT; 2283 AA.  
AC Q8VQ99;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Serine-threonine rich antigen.  
OS Staphylococcus aureus.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1280;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FR1326;
RA Sharp L.J., Henderson B., Poole S., Nair S.;
RT "Identification of a putative serine-threonine rich antigen from
RT Staphylococcus aureus.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF459093; AAL58470.1; -
DR InterPro; IPR008985; Cons like lec_g1.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR008009; He_PIG.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF05345; He_PIG; 2.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
SQ SEQUENCE 2283 AA; 228866 MW; 9C0991E0E59B24B0 CRC64;

Query Match 5.8%; Score 161; DB 2; Length 2283;
Best Local Similarity 20.2%; Pred. No. 0.29;
Matches 117; Conservative 98; Mismatches 263; Indels 102; Gaps 17;

QY 21 ISKRATLDWLSNEATVATPAIINIGADGAWVSGADSGIVASPTDNPDEFTWTRDS 80
DB 748 VTNSNSMSDS-VTSNGSTQSGQSVSTSKADSGASTSGSIVSTKSTKSTSVLSDS 806
QY 81 GLVLTIVDLFRNGDTSLSLTENYISAQIVOGISNPGDLSGAGLGPKNVDETAY 140
DB 807 VSASKSLSTSESNVS--STSTSLVNSQGVSSMSDSA--SKSTLSISNSSTK 861
QY 141 TSGWGRPQRGPALRATAMIGFGQWLNDG-----YTSTA-----TIYWPVINDL 167
DB 862 SEBLSTSTSD-SLRSTSLSDLSMTSGSLSKQSLSTSDASTSGSVSDSTSNST 919
QY 188 SYVAQYVNOTGYDLMEEVNGSFFTTIAVQHRALVEGSAFATAVGSCSCDQAPET--- 244
DB 920 S-----TAEISLSEASSTDSISIS-----NSINSGASSTKSDSGSTSTSL 962
QY 245 -----LCYIQSF-----WTGFTLANFDS-----SRSGKANTLIGSHITPDEA 284
DB 963 TSDSKSMSTSESLSDSTSTSDSVAGSLVAGSQSVSTSTSDSMSTSEIVDSISTGSL 1022
QY 285 ACDDSTQPCSPALANHKVVDVFRSIYTLNDGLSDEAVAVAGRYEDYVNGNPFELC 344
DB 1023 ASBSKSNVSSTSGSTSGSTSLSDSGSTSD--SDSKSLJL-----STSGGS--TS 1072
QY 345 TLAAREQYDALYQMDKQGLLEVTVSLDPEFKALYSDAATGYSSSSSTIVDAVATP 404
DB 1073 TSTSTSSVSTSGSTSGSTSGSTSGSTSTSTSTSTSTSTSTSTSTSTSTSTST 1132
QY 405 ADGFVSLVETHAASNG-----SMSEQDKSDGEQLARDLTWYALLLTANRRNV--- 456
DB 1133 TSGSVSTSTSLSTSNSESTSTSMDSSTSLSTSESDSTSDSTSTSTSTSTSTST 1192
QY 457 -----VPASGERTSASVPG--TCAATSAICTGYSVTVTSPRST 493
DB 1193 LSSGNSGTGDESESSASAFLEBSLSESTSESLSDSTSTSLSDSNSESGSTSTSL 1252
QY 494 VATGTTTATPTGSGSVSTSTKTATATASKTSTTRSGMS 533
DB 1253 NSTSGSTSTSTSTSGSASTSTVKSSESVSTSLSTSTSTSL 1292

RESULT 28
QY7210 PRELIMINARY; PRT; 916 AA.
AC 07Y210;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, last annotation update)
DE MECT1 (Fragment).
OS Monostiga brevicollis.
OC Eukaryota; Choanoflagellida; Codosonigidae; Monostiga.
OX NCBI_TaxId=81824;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 50154;
RX MEDLINE=22752268; PubMed=12869759;
RA King N., Hittinger C.T., Carroll S.B.;
RT "Evolution of key cell signaling and adhesion protein families
RT predates animal origins.";
RL Science 301:361-363(2003).
DR EMBL; AY245772; AAP78680.1; -.
FT NON TER 1
SQ SEQUENCE 916 AA; 97637 MW; 2E9C05CE7FEC6E0 CRC64;

Query Match 5.8%; Score 160.5; DB 5; Length 916;
Best Local Similarity 22.6%; Pred. No. 0.088;
Matches 90; Conservative 48; Mismatches 191; Indels 69; Gaps 13;

QY 175 ATDIWPLV-----RNDLSVAQYVNOTGYDLMEEVNGSFFTTIAVQHRALVGSAP 226
DB 23 ATSTTLVLVAALAGTTEAACNATLNN--GHEYY--YSRDHFHILVN---FTTARAY 75
QY 227 ATAVGSSCWCDSQAPELLCYLQSFWTGFFILANFDSRSKADANTLGSHTFDEAAC 286
DB 76 CRTAGGLVINSKA-----ENDFVPSLNATTFPRRWIGVQYNTSGQDFVNVGVLVD 128
QY 287 DST--FQPCSPALANHKVVDV--FRSIYTLNDGLSDEAVAGRYEDYVNGNPF 340
DB 129 DLITTYTTCDF-----DTCLWFKG--EPNNAASDENCYQGGHNPAPASNPQ 176
QY 341 W-----FLC-----TLAAEQYDALYQMDKQGLLEVTVSLDPEF 375
DB 177 WNDADCTKRAVCEAFSPACMTTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 236
QY 376 KALYSDAATGYSSSSTYSIVDAVATPADGFVSLVETHAASNGMSQYDKSDGEQLS 435
DB 237 TSTSTSTSSSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 294
QY 436 ARDLTSYALTLTANRRNSVVPASGERTSASVPTCATATGICGYSSVTVTSPRSYA 495
DB 295 STSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 354
QY 496 TGCTTTATPTGSGSVSTSTKTATATASKTSTTRSGMS 533
DB 355 T--TTVTSTTTSTSTSTSTSTTTTTLTSTTTTATTS 390

RESULT 29
QY73T2 PRELIMINARY; PRT; 615 AA.
AC 0973T2;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, last annotation update)
DE Hypothetical protein ST0815.
GN ST0815.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxId=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Anka A., Kobugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Yanagi M., Nishimura M., Yamagishi A.,
RA Aoki K.-I., Masuda S., Yanagi M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000983; BAB5828.1; -.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.

```

DR GO:0005976; P:polysaccharide metabolism; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR000165; Glyco\_hydro\_15.  
 DR InterPro: IPR008928; Glyco\_trans\_6hp.  
 DR InterPro: IPR000169; SHPoc\_acsSite.  
 DR Pfam: PF00723; Glyco\_hydro\_15; 1.  
 DR PROSITE: PS00639; THIOI. PROTEASE\_HIS; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 615 AA; 71427 MW; 8A60F9FACB6A99 CRC64;

Query Match 5.8%; Score 160; DB 17; Length 615;  
 Best Local Similarity 21.0%; Pred. No. 0.055;  
 Matches 93; Conservative 56; Mismatches 165; Indels 128; Gaps 20;

QY 37 VARTALNNGADGAVSGADSGIVASPESTNDPFFYTTTRSGVLVLTVDLFRNGDPT 96  
 DB 262 ISLVKKNHVDVNGSIITASSDYFV---GLYDSTNYCPRDSALSAHAL-DIAGYGD 316  
 QY 97 SLSTSTENYISAOAIVOGISNPSGDLSSGAGLGEPRFNVDETAVTGSM-----GRP--- 147  
 DB 317 AMHG--YQYIS-----EVTLEGFLYHKKNPKT-LASSMHWIFRGQIYP 360  
 QY 148 -QRDGPALPAMTIGFGQWLIDNGY-----TSTADIWPLVRNDLSYVAQYWNQT 197  
 DB 361 IOEDETALQV-----WAIANHYOIYKDIIDELIDIFKPFKPAIRFLMRVYEDGLPKP 412  
 QY 198 GVDLMEVNGSFFETAVGHRALVEGSAFATAVSSGSCWDS-----QAPELLCYLQSF 251  
 DB 413 SFLMEERYGHIHYTSTYVAGALVASSELARDIGDEVLASDMLDVAEYKKEVL--RRNV 470  
 QY 252 WTGSEFLANFDSRSRSGDANTLLGSIHTFDEAACDSTFQPCSPALANHEKVD--SF 309  
 DB 471 HNGRFI-----RIDENGN-----KDLVIDASMSF--YFQMVVDRPIM 510  
 QY 310 RSIYTLINDGLSDSEAVAGRPEDTY---YNGNPFELCTLAAEQLYDALYQMDKQSL 365  
 DB 511 NTRALENSIKVGGIA--RYENDMYRKVQPNPILITLMLAEVLD-LGQREK--- 563  
 QY 366 EYTDVSLDFPKALYSDAATCTSSSSSTYSYDAVTFADGVSVIETHAANGSMSEQ 425  
 DB 564 -----ALD-----YINAMSRALPGLPEQ 584  
 QY 426 YDKSGEQLSARDLTWSYALL 447  
 DB 585 VDPENFTSTSVPLVMSHAFT 606

## RESULT 30

Q25336 PRELIMINARY; PRT; 888 AA.

AC Q25336; 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Secreted acid phosphatase 2 (SAP2) precursor.  
 GN LMSAP2.  
 OS Leishmania mexicana.  
 OC Eukaryota; Eukaryozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OC NCBI\_TaxID=5665;  
 RX SEQUENCE FROM N.A.  
 RC STRAIN=MNYC/BZ/62/M379;  
 MEDLINE=95237187; PubMed=7720697;  
 RA Wiese M., Iig T., Lotsepetch F., Overath P.,  
 RT "Ser/Thr-rich repetitive motifs as targets for phosphoglycan  
 modifications in Leishmania mexicana secreted acid phosphatase."  
 RL EMBL J. 14:1067-1074(1995).  
 DR EMBL: Z46970; CA87091.1; -.  
 DR PIR: T46726; T46726.  
 DR GO:0003993; F:acid phosphatase activity; IEA.  
 DR InterPro: IPR000560; HsAc\_phosphatase.  
 DR Pfam: PF00328; acid\_phosphat\_1.  
 DR PROSITE: PS00616; HIS\_Acid\_Phosphat\_1; 1.

KW Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 24 888  
 FT POTENTIAL.  
 FT MATURE SECRETED ACID PHOSPHATASE 2  
 FT (SAP2)  
 SQ SEQUENCE 888 AA; 91354 MW; 3C30B56967755505 CRC64;

Query Match 5.8%; Score 160; DB 5; Length 888;  
 Best Local Similarity 21.1%; Pred. No. 0.091;  
 Matches 116; Conservative 79; Mismatches 257; Indels 172; Gaps 25;

QY 29 SWLSNATYARTAILNNT-----GADGAVSG--AD-----SGI 60  
 DB 154 SWLDNPAAL--HAALNPVIDEHLSDPAIQSAKDAWVEGLCADYNARTNCVLDMDVAA 210  
 QY 61 VVASPESTNP-----DYFTWTRDGLVLTVDLFRNGDPTSLSTI 102  
 DB 211 FEAAGRLDANATIKAVYPEGLQEVNAAMFQVFSWNHTSKDL-TQGSASQNLAQTVLANI 269  
 QY 103 E-----NYISAOAIVOGISNPSGDLSSGAGLGEPRFNV-----DETAVTGSV 144  
 DB 270 NMRILSPSYNMPOYSADHTVTPAVTFPD--QGETMRRPFAVTIFVELLDQTLADASGV 327  
 QY 145 -----GRPDGPALPAMTIGFGQWLIDN-GYTSATDIWPLVRNDLSYVAQY--- 193  
 DB 328 YVALIRGNVVKADGTYYFQESGIKAYCIDEGNKYLAHTGICPL--NSFRRWVDYSRPA 385  
 QY 194 -----MNQGYD-----LMEVNGS-----SFTIAYQHR--- 218  
 DB 386 VADGHCAMTQYQSNDCPRTIADNKPVRSCMLYHVCPSKACPDYILISAVDHQCPG 445  
 QY 219 -----ALVEGSAFATAVSSGSCWCD---SQAPELLCYLQSFMT---GSFLIANF 261  
 DB 446 PDVTNPTSSSSBSGTTSSSEGTATSSDVTTSSSEGTATSSDATTSSSEGTATSSSD 505  
 QY 262 DSGRSGKANDTLGSIHTFDEAACDSTFQPCSPALANHEKVDSPFSIYTLNDGLSD 321  
 DB 506 ATSSSDATTTSSSSSEGT--TSSSDATTTSSSDATTTSSSEGTATSSSDATTTSSSEGTAT 563  
 QY 322 SEAVANGRPEPTLYNNGNWFELCTLAAEQLYDALYQMDKQ---SLEVTDSLDPFKA 377  
 DB 564 SSSDVTTSSSEGT-----ATSSSDATTTSSSEGTSSSDATTTSSSEGTAT 609  
 QY 378 LYSDAATGTYSSSSSTYSYSDAVKTFADGFVSIVETHAANGSMSEQD---KSDSEQ 433  
 DB 610 TSSDATTSS--SSSEGTSSSDATTSSSD---VTTSSSSSEGTATSSSDATTSSSEGTAT 663  
 QY 434 LSARDLTWSYALLTANRRNSVVPASWG-----ETGASSVPGTCAATSAIGTSSVT 486  
 DB 664 TTSSDATTSSSSSEGTSSSDATTSSSEGTATSSSDATTSSSEGTATSSSDATTSSSEGTAT 723  
 QY 487 VTSWPSIVATGTTTATPTGSGSVTSTKTATATKSTTTTS 530  
 DB 724 TATSSDATTSSSSSEGTSSSDATTSSSEGTATSSSDATTSSSDATTSSSDATTSSSDATTSS 767

## RESULT 31

Q990Y4 PRELIMINARY; PRT; 2271 AA.

AC Q990Y4; 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein SAV2654.  
 GN SAV2654 OR SA2447.  
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and  
 OC Staphylococcus aureus (strain N315).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OC NCBI\_TaxID=158878; 158879;  
 RX SEQUENCE FROM N.A.  
 RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);  
 RX MEDLINE=21311952; PubMed=11418146;  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,



RA Cui L., Ogunchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
 RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hosoyma A.,  
 RA Mizutani-U. Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
 RA Sekinizu K., Hirakawa H., Kihara S., Goto S., Yabuzaki J.,  
 RA Kanehisa M., Yamaehita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,  
 RT "Whole genome sequencing of methicillin-resistant *Staphylococcus*  
 RT aureus".  
 RL Lancet 357:1225-1240(2001).  
 DR EMBL: AF003366; BAB58816.1; -.  
 DR EMBL: AP003138; BAB43752.1; -.  
 DR PIR: P90073; P90073.  
 DR InterPro: IPR008985; ConA\_like\_1ec\_g1.  
 DR InterPro: IPR001899; Gram\_pos\_anchor.  
 DR InterPro: IPR008009; He\_PIG.  
 DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam: PF05345; He\_PIG; 2.  
 DR PROSITE: PS00847; GRAM\_POS\_ANCHORING; 1.  
 KM Hypothetical protein, Complete proteome.  
 SQ SEQUENCE 2271 AA; 227844 MW; 7C2A7040D6C8289D CRC64;

Query Match 5.7%; Score 159; DB 16; Length 2271;  
 Best Local Similarity 20.4%; Pred. No. 0.38;  
 Matches 116; Conservative 105; Mismatches 263; Indels 84; Gaps 17;

QY 21 ISKRATLDSWLSNEATVARTALININIGADGAWVSGADGIVASPTDNDYFYWTDRS 80  
 DB 748 VTRNSMDS -VSTSGSTQSGSVSTKADSGASSTSGSIVSTASSTKSTSVLSDS 806  
 QY 81 GLVLTIVDFRNGDTLSTIENYISAQAIQGISNPSGDSGAGLGPKNVDETAY 140  
 DB 807 VSASKSLSTSESNSVS -STSTSLVNSQSV -SSSGSVSKSTSLSDSISNSNTEK 861  
 QY 141 TSGMGRPQRPALRATAMIGFQWLDNGYSTATIIVPL -VRNDLSVAQYWN 195  
 DB 862 SEBLSTSTSD -SLRTSTSLSDSLSMKSTSGSLSGSLSTISGSSSTSLSDSTSNAL 919  
 QY 196 QYGVDMEEVNGSFFTIAYOHRAIVGSAFATAVGSCWCDSQAPRI -LC 246  
 DB 920 STSTSLSESASTSDSIS -NSIANSQASATSKSDQSTSLSTSDSKSMS 970  
 QY 247 YLOSF -WGSFLA -NPDSSRGKANTLLGSIHTFDEACDDSTFQ 292  
 DB 971 TSSSLSDSTSGSVSGSLSLIASQSVSTSDSMSTSEIVSDISITSGSLSDSKSMS 1030  
 QY 293 PCSPRALANKEVDFRSIYTLNDGLDSBAVAGRPEDTYNGNPMFLCTIAAEOL 352  
 DB 1031 VSSSMSTSGSGSTSESLSDSGSTSD -SDSKSL -STSGSGSTSTSTSTSTSTST 1083  
 QY 353 YDALYWDKQGLSEVTVSLDFFKALYSDATGTYSSTSSSTYSIVDAVTFADGFVSI 412  
 DB 1084 SES -QSTSGSMASQSDSMSTSTSPSDSTSDSKASTASSEISQASATSTSGSVST 1140  
 QY 413 ETHAASNG -SMSEQYDKSP -GEQLSARLTYALALTAAN 451  
 DB 1141 TSLSTSESTSTSVSDSTSLSTSESDISESTSTSTSTSTSTSTSTSTSTSTSTST 1198  
 QY 452 RRNSVVAWNGETASASVPGTCATSAIGYSSVTVSWPISVATGTTTAA -PTTGS 508  
 DB 1199 TSSSESGASAFSESLSESTSESTSESTSESTSESTSESTSESTSTSTSTSTSTSGS 1258  
 QY 509 GSV -TSTSTKTATATSKTSTTTTSGMS 533  
 DB 1259 ASISTSTSTSESTSTPFKESVSTSLSMS 1286

RESULT 32  
 Q9URU4 PRELIMINARY; PRT; 1283 AA.  
 AC Q9URU4; Q9USQ3;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Hypothetical protein (SPBC884.07c protein) (SPBC1289.15 protein).  
 GN SPBC1289.15 OR SPBC884.07C.  
 OS Schizosaccharomyces pombe (Pission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OK NCBI\_Taxid=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972h-;  
 RA Wood V., Rajandream M.A., Barrell B.G., Oliver K., Harris D.,  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972h-;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Scouras J., Peat N., Hayles J., Baker S., Baaham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Monev P., Mout R., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Gymnopoulos B.,  
 RA Welfens I., Vanstreels E., Rieger M., Schaefer M., Muller-Auer S.,  
 RA Gabel C., Fuchs M., Dusterhoft A., Fritzc C., Holzer E., Moestl D.,  
 RA Halbert H., Borzym K., Langer I., Beck A., Leirich H., Reinhardt R.,  
 RA Pohl T.M., Eger P., Zimmermann W., Medler H., Wambut R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of *Schizosaccharomyces pombe*.";  
 RL Nature 415:871-880(2002).  
 DR EMBL: AL035675; CAB38695.1; -.  
 DR EMBL: AL021815; CAB6153.2; -.  
 DR GeneDB Spombe; SPBC1289.15; -.  
 DR GO: GO:0016043; P:cell organization and biogenesis; ISS.  
 DR InterPro: IPR003324; BlueCu\_1.  
 DR InterPro: IPR003324; SGXSG.  
 DR Pfam: PF02359; SGXSG; 24.  
 DR PROSITE: PS00196; COPPER\_BLUE; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 1283 AA; 129532 MW; 0B8057CA1C82796 CRC64;

Query Match 5.7%; Score 158; DB 3; Length 1283;  
 Best Local Similarity 21.0%; Pred. No. 0.2;  
 Matches 118; Conservative 76; Mismatches 244; Indels 124; Gaps 24;

QY 5 SLIALSGLVGTGLANVSKRATLDSWLSNEA -TVARTALININIGADGAW -VSGADSGIV- 61  
 DB 180 SIIPHSITSLTSTVSVINDTASLSTKSTSTPTNGITITETIVSGSVGYSTTFPASGTTSTVE 239  
 QY 62 VASPTDNDPFIYTWTRDGLVLTIVDFRNGDTSLSTIENYISAQAIQGISNPSGD 121  
 DB 240 VVEPT -ACTITETIV -SSGVGYSTTFPANGTSGTVEVEPTAG- 281  
 QY 122 LSSGAGLGPKNVDDETATYNGRQRPORPQALRATAMIGFQWLDNGYSTATIDIVMP 181  
 DB 282 -TITETIVSGS -VGYSTFPANGTSGTVEVEPT 313  
 QY 182 LVENDLSVYAQYNGQGYDLMEEVNGSFFTIAYOH -RALVEGSAFAT -AV 230  
 DB 314 TAGTITETIVS -GSVGYSTTFPANGTSGTVEVEPTACTVETITIVSGSVGYSTTFPAS 371

QY 231 GSSCSWCDQAPFILLYLOSFWTGS-FILANPDSSRSGKANTLLGSIHTFDEPAACDS 289  
 DB 372 GTTSGTVEVEPAGTITETIVSGSKAFSTFPAN-----GTTSGTVEVEPAGTITK 425  
 QY 290 TTPQCPALANKKEVVDSPRSITYINDGSDSEAV---AVGYPPEDTYNKNPMLCTL 346  
 DB 426 TIVSGS-----KFTSTFPANGTTSGTVEVEPAGTITE-TIVSGSVGTSFP 473  
 QY 347 AAEEQLYDALYWDKQSLFETDVSIDFFKALYSDAATGTYSSSSSTYSIVDAVTFAD 406  
 DB 474 PA-----NOTTSGTVEVEP-----AGTITETIVSGSKTFTSTFPASGT-TS 515  
 QY 407 GFVSIVETHAASNGSMEQY---DKSDGEQLSARDLTWSYALL--TANNRRNSVVPASW 461  
 DB 516 GTVEVEPEPA---GTITETIVSGSKAFSTFPANGTTSGTVEVEPAGTITETIVSGSV 572  
 QY 462 GEFSSASVETCATSAIGTSSVYT-----SWPVIATGTTTATPTGSG 509  
 DB 573 GYTSFPAGTTSGTVEVEPAGTITETIVSGSVGTSFPASGTSGTVEVEPAGT-AG 631  
 QY 510 SVTSTKTTATAGTSTTTTSG 531  
 DB 632 TITETI-VSGSVGTSFPASG 652

## RESULT 33

OBIFX6 PRELIMINARY; PRT; 2232 AA.

AC 081FX6;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Hypothetical protein K06A9.1b.  
 GN K06A9.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxId=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscot N2;  
 RX MEDLINE=9069613; PubMed=9851916;  
 RA Waterston R.;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscot N2;  
 RA Geisel C., Gattung S.;  
 RT "The sequence of C. elegans coemid K06A9.";  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscot N2;  
 RA Waterston R.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U08046; AAC70890.1; -.  
 DR PIR; T34434; T34434.  
 DR WormPep; K06A9.1b; CB19525.  
 KW Hypothetical protein.  
 SQ SEQUENCE 2232 AA; 213843 MW; 08D69FA638E14C8 CRC64;

Query Match 5.6%; Score 156.5; DB 5; Length 2232;  
 Best Local Similarity 23.4%; Pred. No. 0.54;  
 Matches 113; Conservative 70; Mismatches 228; Indels 71; Gaps 18;

QY 80 SGVLTVLDFNNGTSLSTIENTISAGAIYQGISNPGSDSSGAGLGEPEKNDTERA 139  
 DB 1440 SGGTQHTTWSKASSGT---SPSTNSGTSTVMWSSSTSGVSTSSASSTQPMSTSGS 1496  
 QY 140 YTGSMWRPORDGPALPAMIGFGQWLLDNGYTSATDIWPLVRNDLSVVAQYMNQGY 199

DB 1497 SAGSTVASTAPPAASSTAPSTGT-----MSSTSGTGVSTI-SESSTVASSTQTC- 1548  
 QY 200 DLMEVNGSSFFTIAYO-----HRAIVEGSAFATAVSSCSW--CDQAPILLCYL 248  
 DB 1549 ---STVTMSSSTSGVSTSSASTOPQMSSTQSSAGSTVASTALVSTVPSSTGM 1605  
 QY 249 QSEFWTGSFILANFDSRSGKANTLLGSIHTFDEPAACDSTFPQCPALANKKEVDS 308  
 DB 1606 GSTTSSTGT-VGSTISESTVASASQSGTVMWSSSTSGVSTSSASTOPQMSSTQSSA 1664  
 QY 309 FRSITYTLNDGLDSEAVANGRPEDTYVGNP---WFLCTLAABEQLYDALYWDK----Q 362  
 DB 1665 GSTVASSTTGL-----VSTSTVPSSTGTWGSTSGTSGTISBSSTPASASSTQSGTVM 1719  
 QY 363 GSLEVTDLDFPKALYSDAATG---TVSSSSTYSIVDAVKTADGFVSIVETHAAS 418  
 DB 1720 GSSSTSGVST-----SSASSQOPQMSSTQSSAGSTVSSASTPASAS-----STAPSS 1767  
 QY 419 NSGMSQYDKSDGEQLSARDLTWSYALLTANNRRNSVVPASW-----ETSASSVPGT 472  
 DB 1768 TGTMSSTSGTGVSTWSQ-----SSTAASTSHSTGTVTLTGSSSTSSNQMSSTQSSVGS 1822  
 QY 473 CAATSAIGTSSVYTWSPIYATGTTTATPTGSGSVTSKTTATA-SKTSTTTTSG 531  
 DB 1823 TVASTAGLVSTV---PSTGTWGSTSGT---VGSTISESTVASASQSGTVMG 1876  
 QY 532 MS 533  
 DB 1877 SS 1878

## RESULT 34

093N36 PRELIMINARY; PRT; 1306 AA.

AC 093N36;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Ice protein.  
 GN ICEA.  
 OS Pantoea ananatis (Erwinia uredovora).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Pantoea.  
 OX NCBI\_TaxId=553;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=110;  
 RA Tang C.R., Sun F.Z., Zhao T.C.;  
 RT "Cloning and characterization of an ice gene from Pantoea ananatis  
 RT 110.";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF387802; AA070465.1; -.  
 DR GO; GO:0009279; C:external outer membrane (sensu Gram-negative. . .; IEA.  
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000258; Ice\_nucleatn.  
 DR InterPro; IPR000169; SHprot\_acsite.  
 DR Pfam; PF00818; Ice\_nucleatn; 68.  
 DR PRINTS; PR00327; ICENUCLEATN.  
 DR PROSITE; PS00314; ICE\_NUCLEATION; 47.  
 DR PROSITE; PS00639; THIOLEPROTEASE\_HIS; 2.  
 SQ SEQUENCE 1306 AA; 125852 MW; BC1FF66BEDF1606F CRC64;

Query Match 5.6%; Score 156; DB 2; Length 1306;  
 Best Local Similarity 22.7%; Pred. No. 0.28;  
 Matches 127; Conservative 66; Mismatches 229; Indels 138; Gaps 26;

QY 47 GADGAVSGADGIVASTNDPQFYTWTDGSLVLTVDLFFN-----GPTSL 99  
 DB 299 GADSSLIAGSGTQAGESTGTAGSGTGTAKG-----SDLIAGYSGTGTAGDS-- 350  
 QY 100 STIENTISQAIYQGISNPSG-----DLSG-----AGLGEPEKFN 134

```

Db 351 SLIAGYSTAGEDSLTAGYGSTOTAGKSDLTAGYGSTTAGADSLIAGYGSTOTA 410
QY 135 VDEFAVYTGWGRQDRDPALRATAMIGFQWLLDNGYSTATDVIWPLVRNDLSYVAQY 194
Db 411 GERSTOTAGYGSTQ-----TAQKSD---LTAGYGSTGT-----AGDDSLIA--- 450
QY 195 NQGYDLMEVNGSSFFTIYVQHRALVEGSAFATVAGSSCSCDQAPILCT--LQSF 252
Db 451 ---GYGSTOTAGEDSLTAGYGSTOTAGKSDLTAGYGSTTAGYESSLSIGYGSTOTAG 507
QY 253 TGSFILLNPDSSRSKQANTLL---GSIHFFDEPAACDDSTFQPCSPRALANKEVVD 308
Db 508 YGSLTLTAGYGSTOTAGQESDLITGYGSTST---AGANSSLIAGYGSTOTAGSYNSVLTAG 563
QY 309 FRSIYTLNDGL-----SDSEAVA-VGRYPEDTYNG-NPWLCTLAABEQLY 353
Db 564 YGSLTQTRREGSDLTAGYGSTGTAGSDSLIAGYGSTOTAGSYNSLTAGYGSTOTAG 621
QY 354 DALYQMDKQSLVETDVSLFPKALYSDAATGYSSSSSTYSIVDAVKTFADGFVIVE 413
Db 622 -----SVLTGYGSTTAGADSLIAGYGSTOTAGNSILTA-----GYGS--- 662
QY 414 THAASNGS-MSEQYDKSD-----GEOLSA---RDLTWSYALILTANRRNSV 457
Db 663 TOTAGESSDLTAGYGSTTAGADSLIAGYGSTOTAGSYNSLTAGYGSTOTA--REGSVL 720
QY 458 PASWGETSASVPTCATSAIGTYSVTVTWSPIVATG-GTTTGA-----TPTSGSV 511
Db 721 TTGYGSTTAG-----ADSLIAGYGSTOTAGYNSILTAGYGSTQTRQERSDLTAGYGST 775
QY 512 TSTSKTATASKTSTTRSG 531
Db 776 STAGADSLIAGYGSTOTAG 795

RESULT 35
Q8NUJ3
ID Q8NUJ3 PRELIMINARY; PRT; 2275 AA.
AC Q8NUJ3;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE MM2575 protein.
GN MM2575.
OS Staphylococcus aureus (strain MM2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxId=196620;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naito T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "genome and virulence determinants of high virulence community-
RT acquired MRSA."
RL Lancet 359:1819-1827 (2002).
DR EMBL; AP004831; BAB96440.1; -.
DR InterPro: IPR008985; ConsA1ike Jac g1.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR008009; He_PIG.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR Pfam: PF05345; He_PIG; 2.
DR PROSITE; PS08447; GRAM_POS_ANCHORING; 1.
KM Complete proteome.
SQ SEQUENCE 2275 AA; 228407 MW; E676B6BB9E6CE2E CRC64;

Query Match 5.6%; Score 154.5; DB 16; Length 2275;
Best Local Similarity 20.0%; Pred. No. 0.74;
Matches 110; Conservative 106; Mismatches 235; Indels 99; Gaps 16;
QY 17 LANVISKRATLDSLSNE-ATVARTALINNIG-----ADGAWVGAGSGIVAVSPSTDP 70
Db 803 LBSVSAKSLSTSESVSSSTSTSLVNSQSVSSSSMSDSKSTSLSDSTSNSSSTKES 862

```

```

QY 71 DYFYTWTRDSGLVLTVLVDLFRNGDRLSLSTENYISQAQIVOGISNPGDLSGAGLGE 130
Db 863 ESLSSTSDS---LRTSTSL---SDLSMSSTGSLSKSOSLSTSESS---STSAISLD 913
QY 131 PRNVDEYATVSGWGRQDRDPALRATAMIGFQWLLDNGYSTATDVIWPLVRNDLSYV 190
Db 914 STSNAIST-----ESLSGASTSDSISINSMOSAST 949
QY 191 AQYMNQGYDLMEVNGSSFFTIYVQHRALVEGSAFATVAGSSCSCDQAPILCTYLOS 250
Db 950 SKSDSGSTISLSTDSKSMST-----SESLDSTSTSGVSGSLIAAGS----- 996
QY 251 FWTGSLIANFDSRSRGKQANTLLGSIHFFDEPAACDDSTFQPCSPRALANKEVVD 310
Db 997 -----VSTSTSDMSTSEIVSDISITSGSLASADSKMSVSSSMSTSGSTSESL 1048
QY 311 SITLNDGLSDSEAVAVGRYPEDTYNGNWPFLCTLAABEQLYDALYQMDKQSLVETDV 370
Db 1049 DSQSTSD--SDSKSL---STSGSGSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1098
QY 371 SLDFKALYSDAATGYSSSSSTYSIVDAVKTFADGFVIVEVTHAASNG----- 420
Db 1099 DSMSTSTSPDSTSDSKASTASESISQASSTSGSVTSTSTSTSTSTSTSTSTSTST 1158
QY 421 -----SMSEQYDKSD--GEOLARDLTWSYALILTANRRNSVVPASWGETSASV 469
Db 1159 TSLSTESDISESTSTSDSISSEIASBEST--FISLSESNSTSDSESOASAFLESL 1216
QY 470 PGTCAATSAICGYSSVTVTWSPIVATGTTTGA-----TPTSGSV---TSTKTTTAS 523
Db 1217 ESTSESTSESVSSSTSESTSLSDSTSEGSTSTSLNSSTSGASISTSTSTSTST 1276
QY 524 TSTTRSGMS 533
Db 1277 ESTVSTSLSMS 1286

RESULT 36
Q74851
ID Q74851 PRELIMINARY; PRT; 1131 AA.
AC Q74851;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical serine rich protein.
GN SPC18.01C OR SPC74.07C.
OS Schizosaccharomyces pombe (Pisgion yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxId=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Hilbert H., Duesterhoeft A., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 44-1131 FROM N.A.
RC STRAIN=972H-;
RA Murphy L., Harris D., Lyne M., Rajandream M.A., Barrell B.G.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031907; CA21415.1; -.
DR EMBL; AL031543; CA20837.1; -.
DR PIR; T41144; T41144.
DR GeneDB_Spombe; SPC18.01c; -.
DR InterPro: IPR005556; SUN.
DR Pfam; PF03856; SUN; 1.
SQ SEQUENCE 1131 AA; 116770 MW; 9A3506989CBG2DCG CRC64;

Query Match 5.5%; Score 153; DB 3; Length 1131;
Best Local Similarity 20.4%; Pred. No. 0.35;
Matches 112; Conservative 85; Mismatches 199; Indels 154; Gaps 22;

```

Query 32 SNEATVARTAILNNGADGAMVGADGIVAVSPSTNDPQFYTWTRDGLVLTLDVLF 91  
 Db 560 STEALSNLSSTSSASTSYISPSASSYVANSNDP-----YIS-----QTVSSIT 606  
 QY 92 RNCDDTLSTIENYISAQAIVQGISNPSGDLSSGAGIGEPKENVND-----ET 138  
 Db 607 ASCTTSTSEI-----VSTPASNSNTGSLNGTSSFNVNSGSPSQTTPTSSS 654  
 QY 139 AYVSGNRPRQDPALRATAMIGPGWLDNGTSTATDVLWPLVRNDLSYVAQYNNQTG 198  
 Db 655 SLTGSQSLKETSSPAY-VSSTVSTSSVDSSTVNST-----GSSSDSDQSPGTT 705  
 QY 199 YD-----LWEEV-----NGSFFTIAYQHRALVEGSAFATVGS 232  
 Db 706 YSDPTTITSEVSSILSPTSMQSVSRPSSGDAGFNITPISGSSDGETSGYTISS 765  
 QY 233 SCGWCDSDQAEILCYLQSFMTGSPFILANFDSRSRSGKDANTLGSIH-TPPEACDDSTF 291  
 Db 766 NSSQMSASEPQ-----TAFSSSSSATPTTQSSISTSVSSQSMNSYS 810  
 QY 292 QPCSPALANRHKVDSF-RSITYTLNDGLDSEAVANGRIPEDTYNGNWFCLTAAE 350  
 Db 811 SPISNSVSTSISSIASSTYSIPIS-----SIAS 844  
 QY 351 QLYDA-----LYQMDKQSLFVTDVSLDFKALYSDAATGYSSSSSTYSIVDAVTF 405  
 Db 845 SFPDAGGFTSYNGTARG-----FSSSF--ALANSSESGADVLSTI-----AKPTF- 890  
 QY 406 DGFVSIETHAASNGSSEQYDKSDGEQLARDLTWSYALLTANNRNSVVPASMGETS 465  
 Db 891 -----KFSTNSGTSYSIPSS-----SRNEGTSYSNITVTS--STLKRSLTSVS 937  
 QY 466 -ASSVGTCAATSAIG-----TYSSVTVTSWPSIVATGTTTTPPGSGSVTSTKTAT 520  
 Db 938 TASSYASASSNTLSTPEKTFSSSSTLS-ESISINTNLTVKPESSLSSTTGLTSS 996  
 QY 521 ASKTSTTTS 530  
 Db 997 SSTTPSTTS 1006

RESULT 37  
 ID 08E473 PRELIMINARY; PRT; 1310 AA.  
 AC 08E473;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 GN GBS1529.  
 OS Streptococcus agalactiae (serotype III).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=216495;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MEM316 / Serotype III.  
 RX MEDLINE=2242508; PubMed=12354221;  
 RA Glaeser P., Ruenliok C., Buchrieser C., Chevallier F., Frangeul L.,  
 RA Meadek T., Zouine M., Couve E., Lailou L., Poyart C., Trieu-Cuot P.,  
 RA Kunst F.;  
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing  
 RT invasive neonatal disease.";  
 RT Mol. Microbiol. 45:1499-1513(2002).  
 RL EMBL, AL766851; CAD47188.1; -.  
 DR Segalier, gbs1529; -.  
 DR InterPro: IPR001899; Gram\_pos\_anchor.  
 DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
 DR TIGRfam: TIGR01167; LPTG\_anchor; 1.  
 DR PROSITE: PS00847; GRAM\_POS\_ANCHORING; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 1310 AA; 129397 MM; 9FB7ABBE89B8F155 CRC64;

Query Match 5.5%; Score 153; DB 16; Length 1310;  
 ..Best Local Similarity 20.1%; Pred. No. 0.43; Indels 164; Gaps 19;  
 ..Matches 113; Conservative 75; Mismatches 209

QY 25 ATLDMSLNEATVARTAILNNGADGAMVGADGIVAVSPSTNDPQFYTWTRDGLV 84  
 Db 332 ALKNDTLSEBGVTKLDSENVLVNFAFISAVNEGDFVTKLSDNLD-----TQIGITIL 386  
 QY 85 KTLVDFRNDPDSLSLTIEYISAQAIVQGISNPSGDLSSGAGIGEPKENVND 144  
 Db 387 K-----VODIMETQQLATGSPPTNHI-----TYTW 415  
 QY 145 GRPQDPALRATAMIGPGWLDNGTSTATDI-----VWPLVR-----NDLSYV 190  
 Db 416 TR-----YASTLNINIKARVAMPVPDQRIISKTSDKQCF 451  
 QY 191 AQYNNQTGVDLWEEVNGSFFTIAYQHRALVEGSAFATVGS SCWCDQA----- 241  
 Db 452 ATLNNQVA-SIEERVQYNS-----PSVTEHTNVTNVRSHIMKLDDEROTETVITQI 502  
 QY 242 -PE-ILCYLQSFMTGSPFILANFDSRSRSGKDANTLGSIH-TPPEACDDSTFQPCSPAL 299  
 Db 503 NPEGKEMTPASGLGNLYTIIIGSDGT-SGSPVNLNABVKILK----- 543  
 QY 300 ANRKEVVDSPRSIYT-----LNDG-----LQDSEAVANGRIPEDT 334  
 Db 544 TNSKNLTDSMDQYDSEPEDEDVTQSYNTNDGSKITIDMKTNISSTISYVVLVKIPKQS 603  
 QY 335 --YNGNPFWFLCTIAAEQLYDLYQMDKQSLFVTDVSLDFKALYSDAATGYSSSSS 392  
 Db 604 GVLVS-----TVSDINQYGSKYSGHNTISGDDANAEI--KLSSASASTSAS 654  
 QY 393 TYSSIVDAVKTADGVSIVETHAASNGSSEQYDKSDGEQLARDLTWSYALLTANNR 452  
 Db 655 TSASM-----SASTSASTSASMSASTSASTSASTSASTSASTSASTSASTS 701  
 QY 453 RNSVVPASMGETSASSVPGCATSAIGTYSSTVTSWPSIVATGTTTTPPGSGSVT 512  
 Db 702 STSA--STASASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSAST 759  
 QY 513 STSKTTATASKTSTTTRSGWS 533  
 Db 760 SASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSAST 780

RESULT 38  
 ID 096U11 PRELIMINARY; PRT; 2117 AA.  
 AC 096U11;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN B7A16.095.  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_TaxID=5141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Schulte U., Aign V., Hoheisel J., Brandt P., Fattmann B., Holland R.,  
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
 RA Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA German Neurospora genome project;  
 RL EMBL, AL513445; CAD11404.1; -.  
 DR InterPro: IPR002889; WSC.  
 DR Pfam: PF01822; WSC; 5.  
 DR SMART: SM00321; WSC; 5.  
 KW Hypothetical protein.

SQ SEQUENCE 2117 AA; 216907 MW; 1631362948BEBA32 CRC64;  
Query Match 5.5%; Score 152.5; DB 3; Length 2117;  
Best Local Similarity 20.8%; Pred. No. 0.9;  
Matches 124; Conservative 83; Mismatches 251; Indels 137; Gaps 25;  
QY 29 SWLSNEATVARTALINNIGADGAWSGADGIVVASSTNDPDIYFTWTD---SGVL 84  
DB 798 SVLTGSAWNSOPIALLEHSTSEGGVNT-----IFCNDPPTQCTFENIDMGTSPLTT 848  
QY 85 KTLVLDLFRN-GDPSLSTIENYISAQIVOGISNPSGDLSSGAL-----G 129  
DB 849 TETVTRITNKGSSLM--IDKSKEPMGSLVGAQNPSDDLFEQVNIKKGKESATLFTFP 906  
QY 130 EPEKNVDEATYTSWGRPORBPALRATAMIGFQMLDNGYSTARDI--VWPLVNDL 187  
DB 907 AAPLNADPIYVSGAWTLNVND-----LTFGVHVNFIGLRAIKVGTLPLDSARF 957  
QY 188 SYVAQYNQGYDL-----WEVNGSFEFTTAVGHRALVEGSAFATA-----VGS 232  
DB 958 KYLGCTYDSSANRLTTOAQFPSPDNDNGK-----CQOYAITNKAAPAGTQYTYECWGR 1011  
QY 233 SCGWCDQAPELLC--YL-----QSF--WTGSPILANFDSRSGKANTLLGSIHTFDE 283  
DB 1012 SIPPASLKVDDYLCNTYICREDKSOFCGVSSTMMWMTDTTGYPENGTILAPARPPASK 1071  
QY 284 AACDDSTFOPC-----SPRALANHKEVND-----SFRSI-----YT 314  
DB 1072 AVVGDEWYAGCRTDNASPATRALNDRIYGGSSNTTIESCAQACAGSFCVEVNGCYC 1131  
QY 315 LNDGLSSEAVAGRYEDT---YNGNPWFLLCTLAABQLYDALYQMDKQGSLEVTDS 371  
DB 1132 GNN-LINGSTVA---DEKTCNVVCGDPTELCCGSGRSIVY-----KQKGTVINPS 1179  
QY 372 LDFEALYSDAATGYSSSTVSSIVDAVTFAD-----GFYSIVETHIASNGSMSEQ 425  
DB 1180 TGVSSS--SGTASGTASATASSTSSATASGTPGMPQSIGYSSIGCTSDAVASSLQ 1237  
QY 426 YDKSDGQOLASDLT---WSY-----AALLTANRRNSVVPASNGETSA--SSVPG 471  
DB 1238 GKNTQSVWMSLDLCATYCAQKRYGTEYSACFCGNLLNGAALVDTGRCMCLNGNQOQ 1297  
QY 472 TCATSAIGYSSVTVTSWPSIVATGTTTATPTGSGSVTSKTTATASKIST 526  
DB 1298 ICGSNGISWYQULNPNGTSSSVTASGATOSATASGTASSSSALATTITSS 1352  
RESULT 39  
QY07U3X4 PRELIMINARY; PRT; 2014 AA.  
AC Q7U3X4;  
DT 01-OCT-2003 (TREMBlrel. 25, Created)  
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN SYNW2303.  
OS Synecococcus sp. (strain WH8102).  
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.  
OX NCBI\_Taxid=84588;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=2282697; PubMed=12917641;  
RA Palenik B., Brahmasha B., Larimer F.W., Land M., Hauser L., Chain P.,  
RA Lamerin J., Regala W., Allen E.E., McCarron J., Paulsen I.,  
RA Diferene A., Partensky F., Webb E.A., Waterbury J.;  
RT "The genome of a motile marine Synecococcus.";  
RL Nature 424:1037-1042 (2003).  
DR EMBL; BX569695; CAB08818.1;  
KM Hypothetical protein; Complete proteome.  
SQ SEQUENCE 2014 AA; 215624 MW; 28594D70E8965DBC CRC64;  
Query Match 5.5%; Score 152; DB 16; Length 2014;  
Best Local Similarity 20.2%; Pred. No. 0.91;

Matches 125; Conservative 80; Mismatches 217; Indels 196; Gaps 25;  
QY 28 DSMLSNEATVARTALINNIGADGAWSGADGIVVASPSTNDPDIYFTWTRDGLVTKL 87  
DB 1075 DNALWSEAVNQYTA-----DITDITGKPYLDPENLYL--VQPSGSTRKSA 1118  
QY 88 VDLFRNGDLSLSTIENY-----ISQALVQGISNP 118  
DB 1119 V-LTEGDAFLATESPSYGNPAFSQAPLAVKDVITGSGESALYKLLFEQTITGSSDP 1177  
QY 119 SGLSSGAGLGEKFNVDVETAYTSWGRPORBPALRATAMIGFG-----QWLLD----- 168  
DB 1178 VKRYMT-----LNDQVSMYVD-----ATTTFESHETLEDQRLDND 1218  
QY 169 -----NGYSTARDIWPVLR-----NDLSYVAQYNQGYDL-----WE 203  
DB 1219 GSVFEVANSSTALADTDTAGAKLRQTSDSGLFKDGENSAFVATYADGLRVDDVQETLS 1278  
QY 204 EVNGSFEFTTAVQ-----HRALVEGSAFATAVSSSCWCDQAPELLCYIQSFWTGSL 258  
DB 1279 DVLSRSTKALAVQSGDLYLVKETTGLTVDEVE-----TSVYV 1319  
QY 259 ANFDSRSRG--KQANTLLGSIHTFDEPAACDDSTFOPCSPRALANHKEV---VDSFRSI 312  
DB 1320 VYDVTYTAGIFKSSAL-----YRTBAENETWS-----QDITGDGVVSGGT 1363  
QY 313 YTLNDGLSSEAVAGRYEDTYNGNPWFLLCTLAABQLYDALYQMDKQGSLEVTDSL 372  
DB 1364 SSASDAFADAVGTDVSGEVDVDFONSASDILSVKADG--SNPLSIFVFPSSAKSIVDL 1422  
QY 373 D-----FPKALYSDAATG--TYSSSSSTVSSIVDAVKTFFADGFSIVS----- 413  
DB 1423 TVKQVSDAPLMAKAYDGLTAQVQSRRAKTTDTSYEAVTGLDQITVTNQAEGKI 1482  
QY 414 -----THAANGSMSE---QYKSDGE--QLSARDLTWGYAALLTANR 452  
DB 1483 QSVSWVLPVGVADPTIYAKNSVTGEFTNIQYAEFEGQAMBERTD---LTLFRDNG 1539  
QY 453 RNSVVPASWGETSASVPGTCAATSAIGYSSVTVTSWPSIVATGTTTATPTGSGSVT 512  
DB 1540 KND-----ESDDQGITRRPGVAAASGTTSDSTTPPTPEVPTP--PTPATPTPTPTPT 1593  
QY 513 STEKTTATASKISTTTRS 530  
DB 1594 PTPPTPTPTPATPTPTPS 1611  
RESULT 40  
QY0874K9 PRELIMINARY; PRT; 1383 AA.  
AC Q874K9;  
DT 01-JUN-2003 (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Agglutinin-like protein.  
GN ALS5.  
OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_Taxid=5476;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=SC5314;  
RA Zhao X., Nussens J.A., Hoyer L.L.;  
RT "Analysis of Candida albicans ALS5 function.";  
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY227440; AA072529.1;  
SQ SEQUENCE 1383 AA; 145639 MW; 70A809E8AD79AC41 CRC64;  
Query Match 5.4%; Score 150; DB 3; Length 1383;  
Best Local Similarity 20.5%; Pred. No. 0.72;  
Matches 113; Conservative 84; Mismatches 197; Indels 158; Gaps 26;

QY 26 TLDSWLSNEATVARTAIINNIGADGAMVSGADGIVVASESTDPNPFYPT--WTRDSGLV 83  
 DB 440 TTQFM--SESFTSTTTITNSL-----KGTDS-VIVREP--HNPVTTFTEFSS-----481  
 QY 84 LKTLVULFRNGDLSLSTENYISAQAIYVG-----ISNPSGLSSGAGGEPKFN 135  
 DB 482 -----ESVATETITTTGPGTDSVIIKEPNPTVTITVFWSESYAT 522  
 QY 136 DETAYTSGMCR-----PORDPALRATAMIGFGOMLDNGYSTATDIWPM-----LVR 184  
 DB 523 TEITTTGPGTDSVIIKEPNPTVTITTKF-----W--SESYATETITTTNKEGTDSVIYR 575  
 QY 185 ---NDLSVYAQYWNQGYDLMEVNSSPFTIIVOHRAIVEGSAFATAVGSSCWCDSQA 241  
 DB 576 EPHNPVTTFTEFSES-YATETITTTGPGTDSIIVHDPLEESSSTTAIESSDSN1SSSA 634  
 QY 242 PELTCLTQSWTGSFILANFDSRSGKANDTLGSIHTFPPEACDSTFQPCSPALAN 301  
 DB 635 QE-----SSSVQSSSIVGLSSSDIPLSSD---MPSSS-----666  
 QY 302 HKEVDSFRSITYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABQLYALYQWDX 361  
 DB 667 -----STGLTSSSESTVSYDSSSSSE---LSTFSSSESY-----700  
 QY 362 QGSLFVTVSLDFKALYSPDATGTYSSSSSSTYSIVDAVKTPADGFSIVETHAASNG- 420  
 DB 701 ---SSSISDIT-NFW-----DSSSDLESTISITWSSSIDA-----QSSQSVQSVNSI 744  
 QY 421 SNEQYDKSDGEO--ISARDLTVSYALLTANNRNSVVPASV-----GE---463  
 DB 745 STSQETSSSGEESNTSVTDILVSDASSILNDSISSYPSSTISLSDPEPTIAGEPDS 804  
 QY 464 TSASSVPGTCAATSAIGTYSVTWSPSVATGCTTTATPTGSSSVT-STSKTATAS 522  
 DB 805 RSSSSIASIVETISDLVSLTSDPTSSFDSSSLNSDSSSPFSDSDISASSFSFTLVAP 864  
 QY 523 KTSITTRSGMSL 534  
 DB 865 SFLSSSSSSLSL 876

## RESULT 41

Q08294 PRELIMINARY; PRT; 967 AA.  
 AC 008294;  
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
 DE Chromosome XV reading frame ORF YOL155C.  
 GN YOL155C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_Taxid=4932;  
 RN 1) \_\_\_\_\_  
 RP SEQUENCE FROM N.A.  
 RA Gallion L., Dujon B.  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RN 12) \_\_\_\_\_  
 RP SEQUENCE FROM N.A.  
 RA Arino J., Casamayor A., Gamo F.J., Gancedo C., Lafuente M.J.,  
 RA Aldea M., Casae C., Herrero E.  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RN 13) \_\_\_\_\_  
 RP SEQUENCE FROM N.A.  
 RA MIPS;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; 274897; CAA39177.1; -.  
 DR PIR; S66852; S66852.  
 DR SGD; S0005515; YOL155C.  
 SO SEQUENCE 967 AA; 94704 MW; 7BFC01EA243A551E CRC64;

Query Match 5.4%; Score 149; DB 3; Length 967;

Best Local Similarity 20.4%; Pred. No. 0.51;  
 Matches 117; Conservative 55; Mismatches 201; Indels 200; Gaps 20;  
 QY 10 SGLVCTGLANVSKRATLDSWLSNEATVARTAIINNIGADGAMVSG---ADSGIV-----61  
 DB 434 SGLSLSLESTKGEVTFSPYNSGAFSPSNAILN-----GGSVSLQRRAESGVNNGE 488  
 QY 62 -----VASPSTDN-----PDYFYTWT---RDSGLVLTVDLFRNG 94  
 DB 489 INLENGSTVWVEPVSQSGTINIISGNLYLHPDRTTGQTIVFKGCVL---AVDPTEN 545  
 QY 95 DTSL--STIENYISAQAIYOGIS-----NPSGLSSGAGIGEPKFNDET 138  
 DB 546 TPIPVGYGTGNOALITADYATLALSYDANVGLTATQNSQFSIGTFSSSGFVNSG 605  
 QY 139 ATGWSGRPQROGPALRATAMIGFGOMLDNGYSTATDIWPMVANDLSVYAQYWNQ 198  
 DB 606 TGAAYAVYLYNGVAVSATPS-----STSTTSQATN-----638  
 QY 199 YDLMEVNSSPFTIIVOHRAIVEGSAFATAVGSSCWCDSQAPELICYLQSPWTSFIL 258  
 DB 639 -----STGSGTSPGASV-----TGSTAS 656  
 QY 259 ANFDSRSGKANDTLGSIHTFPPEACDSTFQPCSPALANHKEVDSFRSITYTLNDG 318  
 DB 657 TSFGASVTGSTASTLISG-----SP-----SVYTTTLT 684  
 QY 319 LSDSEAVAVGRYPEDTYNGNPMFLCTLAABQLYALYQWDXGSLFVTVSLDFKAL 378  
 DB 685 YATITSTVAVSCSETTDSNGNVTITTVPCSTTATITSCBETGCHVTST-----736  
 QY 379 YSDAATGTYSSSSSSTYSIVDAVKTPADGFSIVETHAASMSRQYKSDGEOULARD 438  
 DB 737 -GTVAETVSSKS-----YTVYTHCNGNCTVTSSECEESTAT 779  
 QY 439 LT-WSYALLTA---NNRNSVVPASWGETSASVPGTCAATSAIGTYSVTWSPSVIV 494  
 DB 780 TSPKSYTVTVTHCDNGCNT-----KTVISAEATVTTVSPRTYTTATVYTCDD-- 830  
 QY 495 ATGCTTTATPTGSGSVTSTKTATASKTSTT 527  
 DB 831 -NGCSTKVTIVSEAPKETSETSETSA-APKTYTT 861

## RESULT 42

Q05164 PRELIMINARY; PRT; 1001 AA.  
 AC 005164;  
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)  
 DE AOB567, AOF1001, AOE110, AOS264 and AOE130 genes.  
 GN AOF1001.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_Taxid=4932;  
 RN 1) \_\_\_\_\_  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FY1679;  
 RA Gamo F.J., Lafuente M.J., Casamayor A., Aldea M., Casae C., Arino J.,  
 RA Herrero E., Gancedo C.,  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X89715; CAA1860.1; -.  
 SO SEQUENCE 1001 AA; 97673 MW; F11BC4522DCFBDAF CRC64;

Query Match 5.4%; Score 149; DB 3; Length 1001;  
 Best Local Similarity 20.4%; Pred. No. 0.54;  
 Matches 117; Conservative 55; Mismatches 201; Indels 200; Gaps 20;  
 QY 10 SGLVCTGLANVSKRATLDSWLSNEATVARTAIINNIGADGAMVSG---ADSGIV-----61  
 DB 468 SGLSLSLESTKGEVTFSPYNSGAFSPSNAILN-----GGSVGLQRRAESGVNNGE 522

```
QY 62 -----VASPSTDN-----PDYFTWT--RDSGLVLTIVLDFRRG 94
DB 523 INLENGSTYVVEVPVSGSGTINIISGNLYLHPDTFTGQTVFKEGVL--AVDPRETN 579
QY 95 DTSL--STIENVISAQAIYVQGIS-----NPSGDLSSAGLGEKENVDET 138
DB 580 TTP1PVVGYTGEOIALTADVTALSYDSATGVLTAQNSQFSFGTGFSSSGFVNSEG 639
QY 139 AYTGSWGRPQRDPALRATAMIGFGQWLDNGYTSRTADIVMPLVRNDLSYAQYNNQTC 198
DB 640 TPAQAVAYIYANQGVVASSATPS-----STSTTSGATN-----672
QY 199 YDIMEEVNGSSFTLIAVQHRALVEGSAFAPAVAGSSCGWCDSQAPETLCYLOSFWTGSFIL 258
DB 673 -----STGSGTSPGASV-----TGSTAS 690
QY 259 ANFDSSSSGKDANTLLGSIHTFDEACADSTPQPCSPALANHKEVVDSPRSIYTLNDG 318
DB 691 TSPGASVYTGSTASTLISG-----SP-----SVYTTTLT 718
QY 319 LSPSEAVAVGRYPEDTYNGNPMFLCTLAABQLYDALYQMDKQSLVTDVSLDFKAL 378
DB 719 YATTTSTVAVVSCSTTDSNGNVITTTTVCSSSTATITSCDETGCYVTTST-----770
QY 379 YSDAATGYSSSSSTYSSIVDAVKTFPADGFVSIETHAASNGSMSEQYDKSDGEQLSARD 438
DB 771 -GTVAETVSSKS-----YTVTVTHCDNNNGCNTKTVTSECPRETSAT 813
QY 439 LT-WSYAAALLTA---NNRRSVVPASWGEFSASVPTGCAATSAIGYSSVTVTWSISIV 494
DB 814 TSPKSYTVTVTHCDNDGNCMT-----KTVTSEAPATTTTVSPKTYTATVYQCD-- 864
QY 495 ATGCTTTATPTGSGSVTSKTTATASKTSTT 527
DB 865 -NGSTKTVTSEAPKETSETSETSA-APKTYTT 895
```

## RESULT 43

```
070737 PRELIMINARY; PRT; 2230 AA.
ID 070737
AC 070737;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, last sequence update)
DE 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE Conserved hypothetical precursor.
GN SYN00985.
OS Synecchococcus sp. (strain WH8102).
OC Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.
OX NCBI_TaxID=84588;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22825697; PubMed=12917641;
RA Lamerik B., Brahamsa B., Larime F.W., Land M., Hauser L., Chain P.,
RA Lamerik J., Regala W., Allen E.B., McCarren J., Paulsen I.,
RA Dufrene A., Patensky F., Webb E.A., Waterbury J.;
RT "The genome of a motile marine Synecchococcus";
RT Nature 424:1037-1042(2003).
KM EMBL: BX569691; CAB07500.1;
KM Hypothetical protein; Signal; Complete proteome.
FT SIGNAL
SQ SEQUENCE 2230 AA; 226712 MW; C55BA0582586F7E9 CRC64;
```

```
Query Match 5.3%; Score 147; DB 16; Length 2230;
Best Local Similarity 21.2%; Pred. No. 2.2;
Matches 129; Conservative 90; Mismatches 224; Indels 166; Gaps 31;
```

```
QY 1 MSFRLSLALSG-----LVCTGIANYISKRTATD-----SWLSNE 34
DB 690 LAFSTNTSISGSNMASGVLTUTSDSLANAYKALSVTYNNSTSDNPNSTNRISISVWVD 749
QY 35 ATVAKTILNNTGA---DGAWSGADSGIVAVSPSTDNPNFYFTWRDGLVLTIVL 90
```

```
DB 750 GDTNSSAVTSTISVAADVNDAPVTSASATLASEEDGATP-----IDSL--TTTDA 799
QY 91 FRNGDTLSLTEN-----YISAQAIYVQGISNPSGDL-SGAG-LGEKENVDET 138
DB 800 DQSNIESATITISSGFQSTEDVLAFDTSATISQMDASIGVLTLLTSSGTLANYKALESV 859
QY 139 AYTGSWGRPQRDPALRATAMIGFGQWLDNGYTSRTADIVMPLVRNDLSYAQYNNQTC 198
DB 860 TYNNTSDTPTNSNRII-----SMVNDGDTNESA-----VSTISVA-- 897
QY 199 YDIMEEVNGSSFTLIAVQHRALVEGSAFAPAVAGSSCGWCDSQAPETLCYLOSFWTGSFIL 254
DB 898 -----VNDAPVYAGASATLASEGNG-ATIIDSLITTDADSNIESATVISISGFGSA 950
QY 255 SFLIANFDSS-----RSGKDANTLLGSIHTFDEACADSTPQPCSPALANHKEVVD 308
DB 951 EDVLAFSDTSALIGWSNMASGVLTLLGSDSLANYKALFESVTVNNST-----DNPVTAA- 1004
QY 309 FRSI-YTLNDGLSDSEAV-----AVGRYPEDTYNGNPMFLCTLAABQLYDALYQMD 360
DB 1005 -RTISWVNDGDTNSSAVTSTITVTVNNAP--VISGAS--ATLAFSEE--DGATPID 1055
QY 361 KQSLVTDVSLDFKALYSDAATGYSSSSSTYSSIVDAVKTFPADGFVSIETHA-ASN 419
DB 1056 --SSLTITD-----ADDTNIESATVISISGFGSAEDVLAFLDT 1091
QY 420 GSNSEQYDKSDGE-QLSARDLTWSYALL--TAN-----NNRRSVVPASW--GETS 465
DB 1092 SSIISGSNMASGVLTLLGSGTLANYKALESVTYNNSTPTPTNSNRII--SWVNDGDTN 1149
QY 466 ASSVPTGCAATSAIGYSSVTVTWSISIVATGTTTATPTGSGVTSKTTATASKTS 525
DB 1150 SSAN-----TSTISVAADVNDAP--VYAGASATLASEGAGTIDSLTITDADS 1198
QY 526 TTRSGMSL 534
DB 1199 NIESATVISI 1207
```

## RESULT 44

```
097YV0 PRELIMINARY; PRT; 612 AA.
ID 097YV0
AC 097YV0;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, last sequence update)
DE 01-OCT-2001 (Tremblrel. 18, last annotation update)
DE Glucan 1,4 alpha glucosidase (Glucanase) (EC 3.2.1.3).
GN SSO2473.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OX Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332236; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aweyer M.J., Chan-Welher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moers A., Eranua G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.B., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
```

```
EMBL: AE006846; MAK42610.1; -.
FIR: C90419; C90419.
DR GO: GO:0004339; F:glucan 1,4-alpha-glucosidase activity; IEA.
DR GO: GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO: GO:0005976; P:polysaccharide metabolism; IEA.
DR InterPro: IPR000165; Glyco_hydro_15.
DR InterPro: IPR008928; Glyco_trans_6np.
DR Pfam: Pf00723; Glyco_hydro_15; 1.
```



KW Hydrolase; Glycosidase; Complete proteome.  
SQ SEQUENCE 612 AA; 70938 MW; 235FF5JF614CA9C6 CRC64;

Query Match 5.3%; Score 146; DB 17; Length 612;  
Best Local Similarity 21.0%; Pred. No. 0.43; Indels 142; Gaps 23;  
Matches 101; Conservative 66; Mismatches 173;

16 GLANVLSKRLT-----DSWLSN-----BATVARTAIL--NNIGADGAWVSGA 56  
213 GLSDYRKRTPALFRKRVADYWRAMLSKVNDYGEYSILRSLILIOSHQONNGALVAGL 272  
57 DSGIVVASTPNDPFFYTWTRDSGLVLTVDLFRNGDTSLSLTENIYSAQIVQGIS 116  
273 DTDIMKFNDDT---YNYVHRRDAVFCIAL-----ELMGVDRSRQFEFTRLRF 319  
117 NNSGDISGAGLGEPRKFNVDETAYTSGMRP-----QRDPALRATM-IGFGQWLL 167  
320 TINGALFH-----RYTVD--GHFGSTWHPMTLDYLPIDEDETALVYALMWFHFKW-K 369  
168 DNGYSTATDVIWPLVRNDLSVVAQWNO-TG-----YDLMEVNGSSPFTIAVGRALY 221  
370 DVDFIT---YYPWVKGIADFLVNTREKATGLPLSPDMEERIGTHFTTTTIVAGLR 426  
222 EGSAPFATVAGSS--CSWCDQAPEILCYLQSFMTGSFTILANFDSRSRGK-----DAN 271  
427 AAADFARFGEDLDAQKETVADQMRNSLDLFWVGDH-YARTIYMEKGQVHKIDETVDS 485  
272 TLGSGHTDPEACDSTFQPCSPRALANKEVDSFISYTLNGLSDSEAVANGRP 331  
486 ILLAPIFNVLP---MDSRP-----VKNLETVIEKLSYKGLV--RYE 523  
332 EPTYV---NGNWPFLCTLAAEQLYDALYQMDKQSLVETDVSDFPKALYSDAATGY 387  
524 GDQYLGAGNNSNWFSTLWLS-QVYSLMGKDK----- 556  
388 SSSSTSYSSIVDAVKTFADPFVSIVETHAASNGMSQYDKSDGEQLSARDLTWSYALL 447  
557 -----AKKIDWVLSKSLPTGVIPEDQIDND-KPVSYSPLAMSHAEI 598  
448 TA 449  
599 RA 600

## RESULT 45

088789 PRELIMINARY; PRT; 860 AA.  
088789;  
01-JUN-2003 (TReMBLrel. 24, Created)  
01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
DE Muramidase (Putative) (EC 3.2.1.17).  
GN LP\_3093.  
OS Lactobacillus plantarum.  
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
OC Lactobacillus.  
OX NCBI\_TaxId=1590;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCIMB 8826 / WCFSL;  
RX MEDLINE=22480296; PubMed=12566566;  
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,  
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,  
RA Fieze M.W.E.J., Stekema W., Klein Lankhorst R.M., Bron P.A.,  
RA Hofer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,  
RA De Vos W.M., Sijzen R.J.;  
RT "Complete genome sequence of Lactobacillus plantarum WCFSL.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).  
DR EMBL: AL935261; CAD65260.1;  
DR GO: GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
DR GO: GO:0003796; F:glycosylase activity; IEA.  
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO: GO:0016998; P:cell wall catabolism; IEA.

DR GO: GO:0009253; P:peptidoglycan catabolism; IEA.  
DR InterPro: IPR002053; Glyco\_hydro.25.  
DR Pfam: PF01183; Glyco\_hydro.25; 1.  
KW Glycosidase; Hydrolase; Complete proteome.  
SQ SEQUENCE 860 AA; 89631 MW; 4DAC7A298BF90F1C CRC64;

Query Match 5.2%; Score 145; DB 16; Length 860;  
Best Local Similarity 20.1%; Pred. No. 0.79;  
Matches 139; Conservative 85; Mismatches 280; Indels 186; Gaps 28;

7 LALSGVCTGLANVLSKRLTDSWLSNEATV---ARTIILNIGADGAWVSGADSGIV 62  
151 VASSTATTTQVSHTS--STAASASTAVYAVKASSTVSAATATKATKAAVAVAY 208  
63 ASPSTNDPFFYTWTRDSGLVLTVDLFRNGDTSLSLTENIYSAQIVQGI--SNRS 119  
209 TAATTSSTD---VMTIGDTRPRVDVDAVSQSTTQSDP--NKLAAQVKTIVKSTGG 264  
120 GDLSGAGLGEPRKFNVDETAYTSGMRPQRDGPALRATMIGFGQWLLDNG--- 170  
265 TDYTNPALNQAANKAKLANVDYHYATF---STDAKSEATNAGF---LVKNVST 318  
171 -----YSTA---TDIWPVENDLS 188  
319 KYLLFADMEDSSSYANATANIAMFWSLDSFGYKNHGVYTSNSYLDRAVAVTVQGSRV 378  
189 YVAQY-WNOTGYDLMEEVNG-----SFFITAVCHRALVEGSAFATAG 231  
379 WRAPQYPTPSANNLWNTNGAWQFSDTALLPFGSDYGYIDVSDVNGLTEDSAGTNTFV 438  
232 SSCSWCDQAPEILCYLQSFMTGSFTILANFDSRSRGKD-----NTLL 274  
439 TTTSSNDTTSEVT-----TDTSATNTSTSSSTTKKASGWYTFPKTRAIKSAADSAR 491  
275 -----GSIHFDPEAACDSTFQPCSPRALANHK 303  
492 TVGTYSKGNRVYNAEITNGETWLRILYSQSEH-FVXIAAKTTTTPASTSKTYRK 550  
304 EYVDSFRSLYT--LNDGLSDSEAVANGRPEDTTY-----NGNWPFLCTLAAEQLYD 354  
551 NETGYKFTKTTAIGSVSDSAKTLGYTKGTVYNAKVTNGETWLRILYSQGHYV 610  
355 ALYQMDKQSLVETDVSDFPKALYSDAATGYSSSSST--YSSIVDAVK---TFADGFV 409  
611 KIS--GAATSTTTTPATSSSKTV--TKATGYTKFTGTTAIGSVVDSAKTLGYTKGPT 667  
410 SIVETHAASNGMSQYDKSDGEQ---LSARDLTWSYALLTJANNRNSVVPV---SWG 462  
668 VYNAKVTNGQTWLRLYLSYGAQHVKISGAATTTTSSKSTATVPASAKTVAGSTYKPA 727  
463 ETSASVPEPCATSAIGTYSVTVTSMPSIYATGTTTATPTGSG----- 509  
728 KTTAIKSSASDA--STVGTYVGNVTYNAKVTNGQTWLRLYSYGAQHYVVGGAAT 786  
510 -SVTSTSKTTATAS---KTSFTTSGMSL 534  
787 TTSSASKTTAAAGTYTFTKTTNIRTAASL 816

Search completed: June 28, 2004, 07:44:51  
Job time : 53 secs

**THIS PAGE LEFT BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 07:40:02 ; Search time 18 Seconds  
(without alignments)  
1546.748 Million cell updates/sec

Title: US-10-038-723-2  
Perfect score: 2771  
Sequence: 1 MSFRLSLMGLVCTGLANV.....SKTTATASKTSTTSGMGL 534

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2742	99.0	640	1	AMYG_ASPNG
2	2620.5	94.6	639	1	AMYG_ASPSH
3	2588.5	93.4	639	1	AMYG_ASPKA
4	1935	69.8	612	1	AMYG_ASPOR
5	1577	56.9	626	1	AMYG_NEUCR
6	1321	47.7	616	1	AMYG_HORRE
7	781.5	28.2	450	1	AMYG_SCHPO
8	725	26.2	604	1	AMYG_RHIOR
9	671.5	24.2	519	1	AMYG_SACFI
10	664.5	24.0	519	1	AMYG_SACFI
11	641.5	23.2	624	1	AMYG_AXAD
12	597	21.5	549	1	AMYG_YEAST
13	522	18.8	767	1	AMYL_SACDI
14	521	18.8	768	1	AMYL_SACDI
15	221.5	8.0	635	1	YGL0_METJA
16	159.5	5.8	1196	1	ICEV_PSEEX
17	159	5.7	1210	1	ICEN_PSEFL
18	158	5.7	1210	1	ICEN_PSEFL
19	158	5.7	1210	1	ICEN_PSEFL
20	157	5.7	1258	1	ICEN_PSEFL
21	156.5	5.6	1148	1	ICEN_PSEFL
22	155.5	5.6	1148	1	ICEN_PSEFL
23	155	5.6	1148	1	ICEN_PSEFL
24	154.5	5.6	1148	1	ICEN_PSEFL
25	151.5	5.5	1034	1	ICEN_PANAN
26	151.5	5.5	1036	1	ICEN_PANAN
27	151.5	5.5	1036	1	ICEN_PANAN
28	150	5.4	1609	1	FIG2_YEAST
29	147	5.3	1479	1	ALAI_CANAL
30	142	5.1	1419	1	ALAI_CANAL
31	141.5	5.1	1419	1	ALAI_CANAL
32	141	5.1	1802	1	HKRI_YEAST
33	140.5	5.1	1567	1	ICEN_XANCT

34	138.5	5.0	600	1	SP96_DICDI
35	138	5.0	797	1	WGUX_HSVB
36	137	4.9	556	1	WSC3_YEAST
37	136	4.9	542	1	CHT2_RHIOL
38	136	4.9	995	1	Y109_YEAST
39	135.5	4.9	537	1	GUX1_PENJA
40	133.5	4.8	1589	1	PHP_DROME
41	132	4.8	827	1	XANP_XANS2
42	130.5	4.7	1537	1	TEA1_YEAST
43	130.5	4.7	1746	1	TENA_PIG
44	130	4.7	706	1	PLB2_YEAST
45	129.5	4.7	1367	1	AMYL_YEAST
46	129	4.7	1025	1	SLAP_CAUCR
47	128	4.6	1169	1	YK82_YEAST
48	127.5	4.6	776	1	ISOA_PSEXP
49	127.5	4.6	881	1	PRV3_YEAST
50	127.5	4.6	1322	1	YAG3_YEAST
51	127	4.6	497	1	FLIC_ECOLI
52	127	4.6	540	1	CH11_RHIOL
53	127	4.6	1140	1	YMG6_YEAST
54	126.5	4.6	776	1	ISOA_PSEXP
55	126.5	4.6	1861	1	APU_THETU
56	126	4.5	567	1	CH13_CANAL
57	126	4.5	1161	1	DAN4_YEAST
58	125	4.5	525	1	NU62_RAT
59	125	4.5	962	1	GUIN_PSEFL
60	124	4.5	562	1	CHT1_YEAST
61	124	4.5	1616	1	SLAP_BACCI
62	123.5	4.5	5703	1	MUSB_HUMAN
63	123.5	4.5	503	1	WSC2_YEAST
64	123.5	4.5	1075	1	FLOS_YEAST
65	122.5	4.4	1382	1	YRF4_YEAST
66	122.5	4.4	1296	1	YRF1_YEAST
67	122	4.4	510	1	YD10_SCHPO
68	122	4.4	699	1	CH11_BACCI
69	122	4.4	121.5	1	EGT2_YEAST
70	121.5	4.4	3584	1	NBEA_DROME
71	120.5	4.3	1374	1	YMN3_YEAST
72	120.5	4.3	1581	1	YRF2_YEAST
73	120.5	4.3	1859	1	YRF3_YEAST
74	120.5	4.3	1859	1	YRF3_YEAST
75	119.5	4.3	1090	1	YRFE_YEAST
76	119	4.3	542	1	YLM2_SCHPO
77	119	4.3	693	1	TGM3_MOUSE
78	118	4.3	1260	1	ALSI_CANAL
79	117.5	4.2	559	1	GAS1_YEAST
80	117.5	4.2	1758	1	Y1R1_YEAST
81	116.5	4.2	1758	1	Y1R1_YEAST
82	116.5	4.2	542	1	SCWB_YEAST
83	116.5	4.2	1045	1	GUIN_CELFI
84	116	4.2	605	1	WSC4_YEAST
85	116	4.2	1005	1	Y456_CHLTR
86	115.5	4.2	1953	1	B1GA_SALTY
87	115	4.2	412	1	H150_YEAST
88	115	4.2	596	1	MYK7_YEAST
89	114.5	4.1	649	1	YAB3_YEAST
90	114.5	4.1	649	1	YAB3_YEAST
91	114.5	4.1	841	1	YAGX_ECOLI
92	114.5	4.1	914	1	YAGX_ECOLI
93	113.5	4.1	373	1	YLU2_PICAN
94	113.5	4.1	526	1	NU62_MOUSE
95	113.5	4.1	755	1	Y572_CHLPP
96	113	4.1	550	1	FLIC_SHIFL
97	113	4.1	1742	1	GUIN_CANAL
98	113	4.1	2432	1	Y43R_IRV6
99	112.5	4.1	420	1	Y43R_IRV6
100	112.5	4.1	549	1	EPD1_CANAL

## ALIGNMENTS

RESULT 1

AMYG ASPNG STANDARD; PRT; 640 AA.  
ID AMYG ASPNG Q92201; Q99179;  
AC P04064; Q92201; Q99179;  
DT 01-NOV-1986 (Rel. 03, Created)  
DT 01-NOV-1986 (Rel. 03, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Glucoamylase (EC 3.2.1.3) (Glucan 1,4-alpha-  
glucosylase) (1,4-alpha-D-glucan glucohydrolase).  
GN GLAA.  
OS Aspergillus niger, and  
OS Aspergillus awamori.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiiales; Trichocomaceae; mitosporic trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5061, 105351;  
[1]  
RN SEQUENCE FROM N.A.  
RC SPECIES=A.niger.  
RX MEDLINE=84261459; PubMed=6204865;  
RA Boel E., Hansen M.T., Hjort I., Hoegh I., Pihl N.P.;  
RT "Two different types of intervening sequences in the glucoamylase  
RT gene from Aspergillus niger.";  
RL EMO J. 3:1581-1585(1984).  
[2]  
RN SEQUENCE FROM N.A.  
RC SPECIES=A.niger.  
RX MEDLINE=84236105; PubMed=6203744;  
RA Boel E., Hjort I., Svensson B., Norris F., Norris K.E., Pihl N.P.;  
RT "Glucoamylases G1 and G2 from Aspergillus niger are synthesized from  
RT two different but closely related mRNAs.";  
RL EMO J. 3:1097-1102(1984).  
[3]  
RN SEQUENCE OF 25-640, AND COMPARISON OF FORMS G1 AND G2.  
RC SPECIES=A.niger;  
RX MEDLINE=86136085; PubMed=3081341;  
RA Svensson B., Larsen K., Gunnarsson A.;  
RT "Characterization of a glucoamylase G2 from Aspergillus niger.";  
RL Eur. J. Biochem. 154:497-502(1986).  
[4]  
RN SEQUENCE OF 25-640.  
RC SPECIES=A.niger;  
RA Svensson B., Larsen K., Svendsen I., Boel E.;  
RT "The complete amino acid sequence of the glycoprotein, glucoamylase  
RT G1, from Aspergillus niger.";  
RL Carlberg Res. Commun. 48:529-544(1983).  
[5]  
RN SEQUENCE FROM N.A.  
RC SPECIES=A.awamori;  
RX MEDLINE=85085934; PubMed=6440004;  
RA Nunberg J.H., Meade J.H., Cole G., Lawyer F.C., McCabe P.,  
RT "Molecular cloning and characterization of the glucoamylase gene of  
RT Aspergillus awamori.";  
RL Mol. Cell. Biol. 4:2306-2315(1984).  
[6]  
RN REVISIONS.  
RC SPECIES=A.awamori;  
RA Nunberg J.H., Meade J.H., Cole G., Lawyer F.C., McCabe P.,  
RT Schweickart V., Tal R., Wittman V.P., Flatgaard J.E., Innis M.A.;  
RT Submitted (FEB-1985) to the EMBL/GenBank/DBJ databases.  
RN  
RN SEQUENCE OF 1-11 FROM N.A.  
RC SPECIES=A.niger; STRAIN=ATCC 10864;  
RX MEDLINE=91168302; PubMed=2076554;  
RA Fowler T., Berka R.M., Ward M.;  
RT "Regulation of the gla gene of Aspergillus niger.";  
RL Curr. Genet. 18:537-545(1990).  
[8]  
RN CONFORMATION OF O-GLYCOSYLATED REGION.  
RX MEDLINE=92189576; PubMed=1546955;  
RA Williamson G., Belshaw N.J., Williamson M.P.;  
RT "O-glycosylation in Aspergillus glucoamylase. Conformation and role  
RT in binding.";  
RL Biochem. J. 282:423-428(1992).

[9]  
RN ACTIVE SITES, AND MUTAGENESIS.  
RC SPECIES=A.awamori;  
RX MEDLINE=90231978; PubMed=1970434;  
RA Sierks M.R., Ford C., Reilly P.J., Svensson B.;  
RT "Catalytic mechanism of fungal glucoamylase as defined by mutagenesis  
RT of Asp176, Glu179 and Glu180 in the enzyme from Aspergillus  
RT awamori.";  
RL Protein Eng. 3:193-198(1990).  
[10]  
RN MUTAGENESIS OF TRP-144.  
RC SPECIES=A.awamori;  
RX MEDLINE=90046622; PubMed=2510150;  
RA Sierks M.R., Ford C., Reilly P.J., Svensson B.;  
RT "Site-directed mutagenesis at the active site Trp120 of Aspergillus  
RT awamori glucoamylase.";  
RL Protein Eng. 2:621-625(1989).  
[11]  
RN MUTAGENESIS.  
RC SPECIES=A.awamori;  
RX MEDLINE=93165653; PubMed=8433972;  
RA Sierks M.R., Ford C., Reilly P.J., Svensson B.;  
RT "Functional roles and substrate locations of Leu177, Trp178 and Asn182  
RT of Aspergillus awamori glucoamylase determined by site-directed  
RT mutagenesis.";  
RL Protein Eng. 6:75-79(1993).  
[12]  
RN CHARACTERIZATION OF CATALYTIC DOMAIN.  
RX MEDLINE=93277459; PubMed=8503847;  
RA Stoffer B., Frandsen T.P., Busk P.K., Schneider P., Svendsen I.,  
RA Svensson B.;  
RT "Production, purification and characterization of the catalytic  
RT domain of glucoamylase from Aspergillus niger.";  
RL Biochem. J. 292:197-202(1993).  
[13]  
RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 25-495.  
RC SPECIES=A.awamori; STRAIN=Var. X100;  
RX MEDLINE=92406872; PubMed=1527049;  
RA Aleshin A., Golubev A., Firsov L.M., Honzatko R.B.;  
RT "Crystal structure of glucoamylase from Aspergillus awamori var. X100  
RT to 2.2-A resolution.";  
RL J. Biol. Chem. 267:19291-19298(1992).  
[14]  
RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 25-495.  
RC SPECIES=A.awamori; STRAIN=Var. X100;  
RX MEDLINE=94253149; PubMed=8195212;  
RA Aleshin A., Firsov L.M., Honzatko R.B.;  
RT "Refined structure for the complex of acarbose with glucoamylase from  
RT Aspergillus awamori var. X100 to 2.4-A resolution.";  
RL J. Biol. Chem. 269:15631-15639(1994).  
[15]  
RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 25-495.  
RC SPECIES=A.awamori; STRAIN=Var. X100;  
RX MEDLINE=94231577; PubMed=8176747;  
RA Aleshin A., Hoffman C., Firsov L.M., Honzatko R.B.;  
RT "Refined crystal structures of glucoamylase from Aspergillus awamori  
RT var. X100.";  
RL J. Mol. Biol. 238:575-591(1994).  
[16]  
RN STRUCTURE BY NMR OF 533-640.  
RX MEDLINE=96266494; PubMed=8683599;  
RA Sorinachi K., Jacks A.J., Le Gal-Coeffec M.-F., Williamson G.,  
RA Archer D.B., Williamson M.P.;  
RT "Solution structure of the granular starch binding domain of  
RT glucoamylase from Aspergillus niger by nuclear magnetic resonance  
RT spectroscopy.";  
RL J. Mol. Biol. 259:970-987(1996).  
[17]  
RN STRUCTURE BY NMR OF 533-640.  
RX MEDLINE=97341228; PubMed=9195884;  
RA Sorinachi K., Le Gal-Coeffec M.-F., Williamson G., Archer D.B.,  
RA Williamson M.P.;  
RT "Solution structure of the granular starch binding domain of

RT Aspergillus niger glucoamylase bound to beta-cyclodextrin.";  
 RL Structure 5:647-661(1997).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 CC glucose residues successively from non-reducing ends of the chains  
 CC with release of beta-D-glucose.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=G1;  
 CC IsoId=P04064-1; Sequence=Displayed;  
 CC Name=G2;  
 CC IsoId=P04064-2; Sequence=VSP\_000262;  
 CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X00548; CAA25219.1; -;  
 DR EMBL: X00712; CAA25303.1; -;  
 DR EMBL: X00712; CAA25304.1; -;  
 DR EMBL: X02465; AAB59296.1; -;  
 DR EMBL: X02465; AAB59297.1; -;  
 DR EMBL: X56442; CAA39825.1; -;  
 DR PIR: A29166; A29166.  
 DR PIR: A90986; ALASGR.  
 DR PDB: 1ACM; 30-SEP-94.  
 DR PDB: 1GLM; 31-JUL-94.  
 DR PDB: 3GLY; 01-NOV-94.  
 DR PDB: 1DOG; 30-APR-94.  
 DR PDB: 1ACQ; 07-JUL-97.  
 DR PDB: 1ACZ; 07-JUL-97.  
 DR PDB: 1GAI; 17-AUG-96.  
 DR PDB: 1GAI; 17-AUG-96.  
 DR PDB: 1KUL; 11-JUL-96.  
 DR PDB: 1KUM; 11-JUL-96.  
 DR GLYCOSULEDB; P04064; -;  
 DR InterPro: IPR002044; CBD\_4.  
 DR InterPro: IPR008291; Glu-a-glycd\_SBD.  
 DR InterPro: IPR00165; Glyco\_hydro\_15.  
 DR InterPro: IPR008928; Glyco\_trans\_6np.  
 DR Pfam: PF00686; CBM\_20; 1.  
 DR Pfam: PF00723; Glyco\_hydro\_15; 1.  
 DR PIRSF: PIRSF01031; Glu-a-glycd\_SBD; 1.  
 DR PRINTS: PR00736; GLHYDRLASE15.  
 DR ProDom: PD001568; CBD\_4; 1.  
 DR PROSITE: PS00820; GLUCOAMYLASE; 1.  
 DR HydroLase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
 DR Alternative splicing; Signal; 3D-structure.  
 FT SIGNAL 1 18  
 FT PROPEP 19 24  
 FT CHAIN 25 640 GLUCOAMYLASE G1.  
 Query Match 99.0%; Score 2742; DB 1; Length 640;  
 Best Local Similarity 99.6%; Pred. No. 9.8e-174;  
 Matches 528; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 181 PLVRNDLSVYAQYWNQGYDLMEBVNGSSFFTLAVGHRALVGSAPATAVSSCSWCD5Q 240  
 Qy 241 APEILICYLOSFWTGSTITLANPSSSRGKANTLLGSIHFDPDEAACDDSTFOCSPRALA 300  
 Db 241 APEILICYLOSFWTGSTITLANPSSSRGKANTLLGSIHFDPDEAACDDSTFOCSPRALA 300  
 Qy 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABQLDALYQMD 360  
 Db 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABQLDALYQMD 360  
 Qy 361 KQGSLEVTDSLDFFKALYSDATGYSSSSSTYSIVDAVKTFADGFVSIYETHAASNG 420  
 Db 361 KQGSLEVTDSLDFFKALYSDATGYSSSSSTYSIVDAVKTFADGFVSIYETHAASNG 420  
 Qy 421 SMSSEQYDKSDGRLSARDLTWSYAAALLTANNRNSVVPASWGTSASSVPGTCAATSAIG 480  
 Db 421 SMSSEQYDKSDGRLSARDLTWSYAAALLTANNRNSVVPASWGTSASSVPGTCAATSAIG 480  
 Qy 481 TYSSVVTWSPISVATGTTTATPTGSGSVSTSKTTATASKTSTTS 530  
 Db 481 TYSSVVTWSPISVATGTTTATPTGSGSVSTSKTTATASKTSTTS 530  
 RESULT 2  
 AMYG ASPSH STANDARD; PRT; 639 AA.  
 AC P22832;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 01-JUN-1994 (Rel. 29, Last annotation update)  
 DE Glucoamylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)  
 DE (1,4-alpha-D-glucan glucohydrolase).  
 GN GLA.  
 OS Aspergillus shirousamii.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_TaxId=5070;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91182400; Pubmed=1368603;  
 RA Shibuya I., Gomi K., Iimura Y., Takahashi K., Tamura G., Hara S.;  
 RT "Molecular cloning of the glucoamylase gene of Aspergillus shirousamii  
 RT and its expression in Aspergillus oryzae.";  
 RL Agric. Biol. Chem. 54:1905-1914(1990).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 CC glucose residues successively from non-reducing ends of the chains  
 CC with release of beta-D-glucose.  
 CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D10460; BAA01254.1; -;  
 DR HSSP: P04064; 1GAI.  
 DR InterPro: IPR002044; CBD\_4.  
 DR InterPro: IPR008291; Glu-a-glycd\_SBD.  
 DR InterPro: IPR00165; Glyco\_hydro\_15.  
 DR InterPro: IPR008928; Glyco\_trans\_6np.  
 DR Pfam: PF00686; CBM\_20; 1.  
 DR Pfam: PF00723; Glyco\_hydro\_15; 1.  
 DR PIRSF: PIRSF01031; Glu-a-glycd\_SBD; 1.  
 DR PRINTS: PR00736; GLHYDRLASE15.  
 DR ProDom: PD001568; CBD\_4; 1.  
 DR PROSITE: PS00820; GLUCOAMYLASE; 1.  
 DR HydroLase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
 KM Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 24  
 FT SIGNAL 25 640 POTENTIAL.

[illegible]

RT	var. kaachi for localization of the raw-starch-affinity site.", [2]
RL	Agric. Biol. Chem. 53:923-929(1969).
RN	SEQUENCE OF 494-538.
RP	Hayaishida S., Nakahara K., Kuroda K., Miyata T., Iwanaga S.; "Structure of the raw-starch-affinity site on the Aspergillus awamori var. kawachi glucomylase I molecule."
RT	Agric. Biol. Chem. 53:135-141(1969).
RL	-I- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D- glucose residues successively from non-reducing ends of the chains with release of beta-D-glucose. -I- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by European bio-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	EMBL; D00427; BAA00331.1; -. PIR; J00479; J00479.
DR	GlycoSuiteDB: P23176; -. InterPro; IPRO02044; CBD_4. InterPro; IPRO08291; Glu-a-glc5d_SBD. InterPro; IPRO00165; Glyco_hydro_15. InterPro; IPRO08928; Glyco_trans_6hp. Pfam; PF00686; CBM_20; 1. Pfam; PF00723; Glyco_hydro_15; 1. PIRSF; PIRSF001031; Glu-a-glc5d_SBD; 1. PIRINTS; PRO0736; GLHYDRLAS15. DR PRODIME; PD001568; CBD_4; 1. DR PROSITE; PS00820; GLUCOMYLASE; 1. KW Hydroxylase; Glycosidase; Polysaccharide degradation; Glycoprotein; Signal.
KW	Signal.
FT	SIGNAL. 1 18
FT	PROPEP 19 24
FT	CHAIN 25 639
FT	DOMAIN 494 538
FT	BINDING 143 143
FT	ACT_SITE 199 199
FT	ACT_SITE 202 202
FT	ACT_SITE 203 203
FT	DISULFID 233 236
FT	DISULFID 245 472
FT	DISULFID 285 293
FT	CARBOHYD 194 194
FT	CARBOHYD 418 418
FT	CARBOHYD 464 464
FT	CARBOHYD 466 466
FT	CARBOHYD 467 467
FT	CARBOHYD 475 475
FT	CARBOHYD 476 476
FT	CARBOHYD 482 482
FT	CARBOHYD 483 483
FT	CARBOHYD 485 485
FT	CARBOHYD 487 487
FT	CARBOHYD 491 491
FT	CARBOHYD 495 495
FT	CARBOHYD 498 498
FT	CARBOHYD 499 499
FT	CARBOHYD 500 500
FT	CARBOHYD 501 501
FT	CARBOHYD 503 503
FT	CARBOHYD 505 505
FT	CARBOHYD 507 507
FT	CARBOHYD 511 511
FT	CARBOHYD 512 512
FT	CARBOHYD 513 513
FT	CARBOHYD 514 514
FT	CARBOHYD 516 516
FT	CARBOHYD 517 517

```

FT CARBOHYD 519 519 O-LINKED (BY SIMILARITY)
FT CARBOHYD 521 521 O-LINKED (BY SIMILARITY)
FT CARBOHYD 523 523 O-LINKED (BY SIMILARITY)
FT CARBOHYD 524 524 O-LINKED (BY SIMILARITY)
FT CARBOHYD 525 525 O-LINKED (BY SIMILARITY)
FT CARBOHYD 526 526 O-LINKED (BY SIMILARITY)
FT CARBOHYD 527 527 O-LINKED (BY SIMILARITY)
FT CARBOHYD 528 528 O-LINKED (BY SIMILARITY)
FT CARBOHYD 529 529 O-LINKED (BY SIMILARITY)
FT CARBOHYD 530 530 O-LINKED (BY SIMILARITY)
FT CARBOHYD 531 531 O-LINKED (BY SIMILARITY)
FT CARBOHYD 533 533 O-LINKED (BY SIMILARITY)
FT CARBOHYD 534 534 O-LINKED (BY SIMILARITY)
SQ SEQUENCE 639 AA; 68271 MW; E112B31A4D8DDB6 CRC64;

Query Match 93.4%; Score 2588.5; DB 1; Length 639;
Best Local Similarity 93.8%; Pred. No. 1,3e-163;
Matches 497; Conservative 15; Mismatches 17; Indels 1; Gaps 1;

QY 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAILNNIGADGAWYSGADSGI 60
DB 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAILNNIGADGAWYSGADSGI 60
QY 61 VVASPESTNDNDYFYTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYOGISNPSG 120
DB 61 VVASPESTNDNDYFYTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYOGISNPSG 120
QY 121 DLSSGAGLGPKNVDEATATGSGWRPQRDPALRATAMGFQOMLNDNGYSTATDIYW 180
DB 121 DLSSGAGLGPKNVDEATATGSGWRPQRDPALRATAMGFQOMLNDNGYSTATDIYW 180
QY 121 DLSSGAGLGPKNVDEATATGSGWRPQRDPALRATAMGFQOMLNDNGYSTATDIYW 179
DB 121 DLSSGAGLGPKNVDEATATGSGWRPQRDPALRATAMGFQOMLNDNGYSTATDIYW 179
QY 181 PLVRNPLSYAQAQWNOGYDLMEEVNGSSPFT1AVOHRALVEGSAFATAVSSCSMCDSDQ 240
DB 180 PLVRNPLSYAQAQWNOGYDLMEEVNGSSPFT1AVOHRALVEGSAFATAVSSCSMCDSDQ 239
QY 241 APEILCYLOSFWTGSFLANFDSSRSKDNALLGSIHTDPDPAACDSTFOPCSPRALA 300
DB 240 APEILCYLOSFWTGSFLANFDSSRSKDNALLGSIHTDPDPAACDSTFOPCSPRALA 299
QY 301 NHKEVVDPRSITLTDNGSDSEAVAVGRPEPTYTNGNWFCTLAALAAQYDALYQMD 360
DB 300 NHKEVVDPRSITLTDNGSDSEAVAVGRPEPTYTNGNWFCTLAALAAQYDALYQMD 359
QY 361 KQGSLEVTDVSLDFPALYSDAATGYSSSSSYSSIVDAVKTGFADGFVSVETHAASNG 420
DB 360 KQGSLEVTDVSLDFPALYSDAATGYSSSSSYSSIVDAVKTGFADGFVSVETHAASNG 419
QY 421 SMSBOYDKSDGQSLARDLTWSYAALLTANRRNSVVPASWGTSAASVPGTCAATSAIG 480
DB 420 SISEQFDKSDGDELARDLTWSYAALLTANRRNSVVPASWGTSAASVPGTCAATSAIG 479
QY 481 TYSSVTVTWSPIVATGCTTTTATPTGSGSVTSTKTTATAGTSTTTTS 530
DB 480 TYSSVTVTWSPIVATGCTTTTATPTGSGSVTSTKTTATAGTSTTTTS 529

RESULT 4
AMYG ASPOR STANDARD; PRT; 612 AA.
ID AMYG ASPOR STANDARD; PRT; 612 AA.
AC P36914;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Glucoamylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE (1,4-alpha-D-glucan glucohydrolase).
DE GLUA.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OK NCBI_TaxID=5062;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=91254744; PubMed=1368680;

```

```

RA Hara Y., Kitamoto K., Gomi K., Kumagai C., Tamura G., Hara S.;
RT "The glucoamylase cDNA from Aspergillus oryzae: its cloning,
RT nucleotide sequence, and expression in Saccharomyces cerevisiae.";
RL Agric. Biol. Chem. 55:941-949(1991).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=92104497; PubMed=1761224;
RA Hara Y., Tsuchiya K., Kitamoto K., Gomi K., Kumagai C., Tamura G.,
RA Hara S.;
RT "Nucleotide sequence and expression of the glucoamylase-encoding gene
RT (glua) from Aspergillus oryzae.";
RL Gene 108:145-150(1991).
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=RIB 40;
RA Hara S., Tamura G., Kumagai C., Gomi K., Kitamoto K., Tsuchiya K.,
RA Hara Y.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D01035; BA00841.1; -
DR EMBL: D10598; BA01540.1; -
DR PIR: JQ1346; JQ1346.
DR HSSP: P04064; IGAL.
DR InterPro: IPR002044; CBD_4.
DR InterPro: IPR008291; Glu-a-glycd_SBD.
DR InterPro: IPR001015; Glyco_hydro_15.
DR InterPro: IPR008928; Glyco_citane_6hp.
DR Pfam: PF00686; CBM_20; 1.
DR Pfam: PF00723; Glyco_hydro_15; 1.
DR PIRSF: PIRSF01031; Glu-a-glycd_SBD; 1.
DR PRINTS: PRO0736; GLHYDRASE15.
DR PRODOM: PD001568; CBD_4; 1.
DR PROSITE: PS00820; GLUCOAMYLASE; 1.
DR HydroLase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT PROPEP 20 25 BY SIMILARITY.
FT CHAIN 26 612 GLUCOAMYLASE.
FT BINDING 146 146 SUBSTRATE (BY SIMILARITY).
FT ACT_SITE 202 202 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 205 205 GENERAL ACID CATALYST (BY SIMILARITY).
FT ACT_SITE 206 206 INTERACT WITH SUBSTRATES (BY SIMILARITY).
FT DISULFID 236 239 BY SIMILARITY.
FT DISULFID 248 475 BY SIMILARITY.
FT DISULFID 288 296 BY SIMILARITY.
FT CARBOHYD 39 39 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 612 AA; 65486 MW; CD7B23B5FA978F97 CRC64;

Query Match 69.8%; Score 1935; DB 1; Length 612;
Best Local Similarity 70.3%; Pred. No. 1,9e-120;
Matches 367; Conservative 59; Mismatches 90; Indels 6; Gaps 4;

QY 1 MSFRSLALSGVCTGLA--NVISKRATLDSWLSNEATVARTAILNNIGADGAWYSGADS 58
DB 2 VFSSCLRALALGSSVLAQVPLRQATGADTLSTEANFSROAILNNIGADGASAGASP 61
QY 59 GIVVASPESTNDNDYFYTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYOGISNPS 118
DB 62 GVIVASPESTNDNDYFYTWTRDSGLVLTVDLFRNGDADLPLIEPFISSOARIQGISNP 121
QY 119 SGLDSSGAGLGPKNVDEATATGSGWRPQRDPALRATAMGFQOMLNDNGYSTATDI 178

```



Db 122 SGALSSS-GAGEPFVNDFTAWGRPDRDGLATMTISFGEMLVENSHTSIATDL 180  
Qy 179 VMLPVRNDLSYVAQYMNQGYDLMEEVNGSSFFITIAVQHRALYEGSAFATAVSSCSMCD 238  
Db 181 VMLPVRNDLSYVAQYMNQGYDLMEEVNGSSFFITIAVQHRALYEGSAFATAVSSCSMCD 240  
Qy 239 SOAPEILICYQSFWTGSFIIANFDSRSGKDANTLGSITHTFDEPAACDSTFQPCSPRA 298  
Db 241 SOAPEILICYQSFWTGSFIIANFDSRSGKDANTLGSITHTFDEPAACDSTFQPCSPRA 300  
Qy 299 LANHKEVVDSPRSIYTLNDGLSDSEAVAGRYPEPTYNGNPNFICLTAAAEOLYDALYQ 358  
Db 301 LANHKEVVDSPRSIYTLNDGLSDSEAVAGRYPEPTYNGNPNFICLTAAAEOLYDALYQ 360  
Qy 359 MDKGSLEVTVDVSDIFFKALYSDAATGTSSSSSTYSIYDAVKTFADEFSIYETHAAS 418  
Db 361 MDKGSLEVTVDVSDIFFKALYSDAATGTSSSSSTYSIYDAVKTFADEFSIYETHAAS 420  
Qy 419 NGSMSEQDYKSDGEQLSARDLTWMSYALLTANNRRNSVVPASWGETSASSVPGTCAATSA 478  
Db 421 TGSMAEQYTKDGSQTSARDLTWMSYALLTANNRRNSVVPASWGETSASSVPGTCAATSA 480  
Qy 479 IGTYSVTVTSWPSIVATGTF--TTTATPTGSSSVTSTSKTT 518  
Db 481 IGTYSVTVTSWPSIVATGTF--TTTATPTGSSSVTSTSKTT 521

RESULT 5

AMYG\_NEUCR STANDARD; PRT; 626 AA.

AC P14804; Q9P5U5; 14, Created)  
DT 01-APR-1990 (Ref. 14, Last sequence update)  
DT 10-OCT-2003 (Ref. 42, Last annotation update)  
DE 10-OCT-2003 (Ref. 42, Last annotation update)  
DE Glucosylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)  
DE (1,4-alpha-D-glucan glucosylase).  
GN GLA-1 OR B5022.70.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxId=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=74-OR23-1A / FGSC 987;  
RX MEDLINE=94037144; PubMed=8221928;  
RA Stone P.J., Makoff A.U., Parish J.H., Radford A.;  
RT "Cloning and sequence analysis of the glucanase gene of Neurospora  
crassa".  
RL Curr. Genet. 24:205-211 (1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=74-OR23-1A / FGSC 987;  
RX MEDLINE=2542210; PubMed=12655011;  
RA Mannhaupt G., Montone C., Haase D., Mewes H.-W., Aign V.,  
RA Honeisel J.D., Fartmann B., Nyakatura G., Kempen F., Maier J.,  
RA Schulte U.;  
RT "What's in the genome of a filamentous fungus? Analysis of the  
RT Neurospora genome sequence".  
RL Nucleic Acids Res. 31:1944-1954 (2003).

RN [3]  
RP SEQUENCE OF 36-65.  
RC STRAIN=74-OR23-1A / FGSC 987;  
RA Koh-Luar S.I., Parish J.H., Bleaaby A.U., Pappin D.J.C., Atanley K.,  
RA Johansen F.E., Radford A.;  
RT "Exported proteins of Neurospora crassa: 1-glucosylase";  
RL Enzyme Microb. Technol. 11:692-695 (1989).  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
CC glucose residues successively from non-reducing ends of the chains  
CC with release of beta-D-glucose.  
CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X67291; CAA47707.1; -.  
DR EMBL: AL355932; CAB91426.1; -.  
DR HSSP: P04064; 1GAI.  
DR InterPro: IPR002044; CBD\_4.  
DR InterPro: IPR008291; Glu-4-glycosidase SHD.  
DR InterPro: IPR000165; Glyco\_hydro\_15.  
DR InterPro: IPR008928; Glyco\_trans\_6np.  
DR Pfam: PF00686; CBM\_20; 1.  
DR Pfam: PF00723; Glyco\_hydro\_15; 1.  
DR PIRSF: PIRSF001031; Glu-4-glycosidase SHD.  
DR PRINTS: PR00735; GLHYDRASE15.  
DR ProDom: PD001568; CBD\_4; 1.  
DR PROSITE: PS00820; GLUCOMYLASE; 1.  
KM Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
KM Signal.  
FT SIGNAL. 1 19  
FT PROPEP 20 35  
FT CHAIN 36 626  
FT BINDING 155 155  
FT ACT SITE 211 211  
FT ACT SITE 214 214  
FT ACT SITE 215 215  
FT CARBOHYD 106 106  
FT CARBOHYD 206 206  
FT CARBOHYD 217 217  
FT CONFLICT 82 82  
FT CONFLICT 550 550  
FT CONFLICT 560 560  
SQ SEQUENCE 626 AA; 66474 MW; 54E5BDBA7A3E349 CRC64;  
V -> L (IN REF. 1).  
A -> R (IN REF. 1).  
MISSING (IN REF. 1).  
N-LINKED (GLCNAC. .) (POTENTIAL).  
N-LINKED (GLCNAC. .) (POTENTIAL).  
N-LINKED (GLCNAC. .) (POTENTIAL).  
SUBSTRATE (BY SIMILARITY).  
CATALYTIC BASE (BY SIMILARITY).  
GENERAL ACID CATALYST (BY SIMILARITY).  
INTERACT WITH SUBSTRATE (BY SIMILARITY).  
N-LINKED (GLCNAC. .) (POTENTIAL).  
N-LINKED (GLCNAC. .) (POTENTIAL).  
N-LINKED (GLCNAC. .) (POTENTIAL).  
MISSING (IN REF. 1).  
A -> R (IN REF. 1).  
V -> L (IN REF. 1).

Query Match 56.9%; Score 1577; DB 1; Length 626;  
Best Local Similarity 57.6%; Pred. No. 8, 6e-97;  
Matches 306; Conservative 85; Mismatches 130; Indels 10; Gaps 6;

Qy 2 SFRSLALSLGLVCTGLANVSKRATLDSMUSNATVARTALININIGADGAMWGASGIV 61  
Db 13 AFOAVLGLPDLPEKHKHSDIKR-SVDSYIQETTPYAKNLLCNIGASGRASGAASGV 71  
Qy 62 VASPSYDNPDPFYTWTRDSGLVKTIVDLFRNG-DTSLSTTENYISAQIVGSIENPSG 120  
Db 72 VASPSYDNPDPFYTWTRDSGLVKTIVDLFRNG-DTSLSTTENYISAQIVGSIENPSG 131  
Qy 121 DLSSGAGLGEPRKVVDELTAVTSGWGRPQRDGPALRTAMIGCOMLIDNGYSTATDIW 180  
Db 132 SLNSGAGLGEPRKFMVDLQOFTGAMGRPQRDGPALRTAMIGCOMLIDNGYSTATDIW 191  
Qy 181 ELYVRNRLSYVAQYMNQGYDLMEEVNGSSFFITIAVQHRALYEGSAFATAVSSCSMCDQ 240  
Db 192 ELYVRNRLSYVAQYMNQGYDLMEEVNGSSFFITIAVQHRALYEGSAFATAVSSCSMCDQ 251  
Qy 241 APRILICYQSFWTGS-FIIANFDSRSGKDANTLGSITHTFDEPAACDSTFQPCSPRAL 299  
Db 252 APRILICYQSFWTGS-FIIANFDSRSGKDANTLGSITHTFDEPAACDSTFQPCSPRAL 311  
Qy 300 ANHKEVVDSPRSIYTLNDGLSDSEAVAGRYPEPTYNGNPNFICLTAAAEOLYDALYQ 359  
Db 312 ANHKEVVDSPRSIYTLNDGLSDSEAVAGRYPEPTYNGNPNFICLTAAAEOLYDALYQ 370  
Qy 360 DKSGSLEVTVDVSDIFFKALYSDAATGTSSSSSTYSIYDAVKTFADEFSIYETHAAS 419  
Db 371 DKSGSLEVTVDVSDIFFKALYSDAATGTSSSSSTYSIYDAVKTFADEFSIYETHAAS 430  
Qy 420 GMSSEQDYKSDGEQLSARDLTWMSYALLTANNRRNSVVPASWGETSASSVPGTCAATSA 479  
Db 431 GMSSEQDYKSDGEQLSARDLTWMSYALLTANNRRNSVVPASWGETSASSVPGTCAATSA 490

QY 480 GYSSVTVTSWPSIVATGTTTT-ATPTSGS-----VTSSTKTATASKT 524  
 DB 491 GSATATATATSFPAHLTPASTTTPPTGTGCAADHEVLVTNEKVTTSYGOT 541

RESULT 6  
 ID AMYG\_HORRE STANDARD; PRT; 616 AA.  
 AC 003045;  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Glucoamylase P precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)  
 DE (1,4-alpha-D-glucan glucohydrolase).  
 GN GAMP.  
 OS Hormoonis resiniae (Creosote fungus) (Amorphotheca resiniae).  
 OC Eukaryota; Fungi; Ascomycota; Ascomycota incertae sedis;  
 OC Amorphothecaceae; Amorphotheca.  
 OX NCBI\_TaxId=5101;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 20495;  
 RX MEDLINE=93146382; PubMed=1490604;  
 RA Joutejoki V.V., Torkkeli T.K.;  
 RT "Glucoamylase P gene of Hormoonis resiniae: molecular cloning,  
 RT sequencing and introduction into Trichoderma reesei.";  
 RL FEMS Microbiol. Lett. 78:237-243 (1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 20495;  
 RX MEDLINE=93365035; PubMed=8358830;  
 RA Vainio A.E.I., Torkkeli H.T., Tuusa T., Aho S.A., Fagerstrom B.R.,  
 RA Korhola M.P.;  
 RT "Cloning and expression of Hormoonis resiniae glucoamylase P cDNA in  
 RT Saccharomyces cerevisiae.";  
 RL Curr. Genet. 24:38-44 (1993).  
 RN [3]  
 RP SEQUENCE OF 72-76, AND CHARACTERIZATION.  
 RX MEDLINE=90338987; PubMed=2116499;  
 RA Fagerstrom R., Vainio A.E.I., Suoranta K., Pakula T., Kalkkinen N.,  
 RA Torkkeli H.T.;  
 RT "Comparison of two glucoamylases from Hormoonis resiniae.";  
 RL J. Gen. Microbiol. 136:913-920 (1990).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 CC glucose residues successively from non-reducing ends of the chains  
 CC with release of beta-D-glucose.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: X68143; CAA48243.1; -;  
 DR EMBL: X67708; CAA47945.1; -;  
 DR PIR: S33908; S33908.  
 DR HSSP: P04064; 1GAI.  
 DR InterPro: IPR002044; CBD\_4.  
 DR InterPro: IPR008291; Glu-a-glcSD\_SBD.  
 DR InterPro: IPR000165; Glyco\_hydro\_15.  
 DR InterPro: IPR008928; Glyco\_trans\_6np.  
 DR Pfam: PF00686; CBM\_20; 1.  
 DR Pfam: PF00723; Glyco\_hydro\_15; 1.  
 DR PIRSF: PIRSF001031; Glu-a-glcSD\_SBD; 1.  
 DR PRINTS: PR00736; GLHYDRLASES.  
 DR ProDom: PD001568; CBD\_4; 1.  
 DR PROSITE: PS00820; GLUCOAMYLASE; FALSE NEG.  
 KW Hydroxylase; Glycosidase; Polyaccharide degradation; Glycoprotein;  
 KW Signal.

FT SIGNAL 1 29  
 FT CHAIN: 30 616  
 FT BINDING 149 149  
 FT ACT SITE 205 205  
 FT ACT\_SITE 208 208  
 FT ACT\_SITE 209 209  
 FT CARBOHYD 200 200  
 FT CARBOHYD 427 427  
 SQ SEQUENCE 616 AA; 66433 MW; B5FADCEBBB152FB CRC64;  
 Query Match 47.7%; Score 1321; DB 1; Length 616;  
 Best Local Similarity 48.6%; Pred. No. 6, 8e-80;  
 Matches 257; Conservative 88; Mismatches 150; Indels 34; Gaps 6;  
 QY 14 CTGLANVIS-----KRALDLSWLSNEATVATATLNNINAGDAWVGSDGIVVA 63  
 DB 8 CAGASLCSLIAIAFTELKARDLSFISERAIALQGLANNIGPDSVPGAGAFVVA 67  
 QY 64 SPSTDNPDYFYTWTRDSGLVLTVDLFRNGDTSLSTIENYISAQAIYVQGISNPSGD-L 122  
 DB 68 SPSKANPDYFYTWSRDSALTLKMIIDEFLGNTTQTITEQYIHAQAVIQTVSNPSGTFL 127  
 QY 123 SSAGLGEPRFVNDETATYGSWGRPDGPALRATMIGFGOMLDNGYTTSTATDIWPL 182  
 DB 128 PPGVGIGBEPKFMVDGTRFNGPWRPQRDPGPAIRALIMTYSMWLIRNGOFAEAKTIWPI 187  
 QY 183 VNRDLSVYVQVNNQGTGYDLMEEVNCSFPTIAVQHRALVSGAFATAVSGSCSWCQAP 242  
 DB 188 IANDLSYVQYNNQSGFDLMEETVYASSPFTIONQHRALVSGAQLADLGTTCGQ-QAP 246  
 QY 243 ELICLYOSFWTSGFILANF--DSRSRGKANTLLGSIHTFDEPAACDSTFPQSPRALA 300  
 DB 247 EVLCFLQSFWMNGKXIVSNINNVNNGRGDNGSLIGSIFTDIDAYCDSPTLQCHSQSLA 306  
 QY 301 NHKEVVDSPRSYITLNDGLSDSEAVAGVRYPEDTYNGNPWFCLTAAABQLYDALYQND 360  
 DB 307 NFKVLTDTPRNLTYTINAGIPEGQGVAVGRVYAEVYVGNPNWYLITAAAEFLYDAVAQWK 366  
 QY 361 KQGSLEVTDVSLDFPFAALSDAATGYSS--SSSTYSIIVDAVKTFADPVSIVETHAAS 418  
 DB 367 ARHVLTVDSTLAFFDIYPEVTVREYKSGNANSPPAQIMDAVTAYADSVYALAEKYIPS 426  
 QY 419 NGSMSEQYKSDQEOQLARDLTMSYVALTANNRRRSVPASGETSASSVPGTCATSA 478  
 DB 427 NGSLSQPRKRDGTPLSALIDLMTSYAFLTMSQKRAQGVPSMSGSNALPPTTCASST 486  
 QY 479 IGTYSSTVTSWPSIVATGTTTTATPTGSGVTSYSTKTATASKTSTT 527  
 DB 487 PGTY-----TPATAGAPNVTSSQVSIFFINATTT 517

RESULT 7  
 ID AMYG\_SCHPO STANDARD; PRT; 450 AA.  
 AC 060087; O96WS5;  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Probable glucoamylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-  
 DE glucosidase) (1,4-alpha-D-glucan glucohydrolase) (meliotic expression  
 DE upregulated protein 17).  
 GN MEU17 OR SPBC14C8.05C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxId=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N.T., Hayles S., Baker S., Baahman D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.C.,  
RA Mooney P., Moulis S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volkart G., Aert R., Roben J., Gymnopoulos B.,  
RA Weltsch I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gebel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Beger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreesen S., Gloux S., Lelaire V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rocher M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.,  
RT "The genome sequence of Schizosaccharomyces pombe.";  
RT Nature 415:871-880(2002).  
[2]  
RN SEQUENCE OF 210-450 FROM N.A.  
RP STRAIN=CD16-1;  
RC MEDLINE=21270454; PubMed=11376151;  
RA Matsubara T., Miyashita K., Saito T.T., Yoneki T., Kakihara Y.,  
RA Nishihama K., Kishi Y.A., Shimoda C., Nojima H.,  
RT "Comprehensive isolation of meiosis-specific genes identifies novel  
RT proteins and unusual non-coding transcripts in Schizosaccharomyces  
RT pombe.";  
RT Nucleic Acids Res. 29:2327-2337(2001).  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
CC glucose residues successively from non-reducing ends of the chains  
CC with release of beta-D-glucose.  
CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: AL023305; CA18423.1; -  
CC EMBL: AB054301; BAB60870.1; -  
CC F1R; T39433; T39433.  
CC HSP; P04064; IGAI.  
CC GeneDB Spombe; SPBC14C8.05c; -  
CC InterPro: IPR000165; Glyco\_hydro\_15.  
CC InterPro: IPR008928; Glyco\_trans\_6hp.  
CC Pfam; PF00723; Glyco\_hydro\_15; 1.  
CC PRINTS; PR00736; GLYHDLASE15.  
CC Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
CC Signal; Meiosis.  
FT SIGNAL 1 16  
FT PROPEP 17 28  
FT CHAIN 29 450  
FT BINDING 147 147  
FT ACT\_SITE 203 203  
FT ACT\_SITE 206 206  
FT ACT\_SITE 207 207  
FT CARBOHYD 383 383  
FT CARBOHYD 409 409  
SQ SEQUENCE 450 AA; 51163 MW; 31C5BFA28E4785FE CRC64;

Query Match 28.2%; Score 781.5; DB 1; Length 450;  
Best Local Similarity 37.6%; Pred. No. 1.9e-44;  
Matches 179; Conservative 68; Mismatches 166; Indels 63; Gaps 12;

QY 7 LALSLVCTG---LANVSKRATLDSWLSNENATVARTAILNINIGADGANVSGANDSIVVA 63  
Db 8 LILGGVAVSASLSPKPKRKEKASMDWTQOQKIAIGHNIAIGSGMAKADINPCIIA 67  
QY 64 SPSTNDPQVFWYMTWRSGLVLTAVDLFRNGPTSLSTLENTIYSAQAIYOGISNPGSDLS 123  
Db 68 SPSTSDPVDYQWVRSRSLATITLTLFRFEGDKLEPIIVKYNDEVRLOAKKPNPBGDPR 127  
QY 124 SGAGLSEPFENVDETAYTSGWGRPQRDGPALRATAMIGCGWMLDNGYSTATDIWPLV 183  
Db 128 AG-GLEPFENVDGTYDDDMGRPQNDSPALRALIATIKMYNLFENKGEVHEVTWIEAV 186  
QY 184 RNDLSVAQYANOTGYDLMEEVNGSSFTIAYOHRVLVGSAPATLVGSSCSMCSDQAP 243  
Db 187 LADLDYATNHWTEASFDLWEEIWDVYFTLAVOKRAMOGTAFKRG-----ADP 237  
QY 244 -----ILCYQSFV-TGSPFLANFDS--SSGQDANTLGSHT--FDPEACD 287  
Db 238 QALVYRTTEPIDLKLGEFMDPOMGVIKGKGVDSGLDSCSTLISLVSNEFD----- 291  
QY 288 DSTFQPCSPRALANHKEVVDSPRSIYTLNDGLSDSEAVAVGRYPEDTYN-----GNPWF 342  
Db 292 -----WHILPTLLKQETMTRDYPVNGMKQ-----AMGRPEVDYGVSGISGNPWF 339  
QY 343 LCTLAASEQYDALYQMDKQGLSEVTDVSLDPFK--ALYSDAATGTYSSSSSTYSIVA 400  
Db 340 ICTSSAEITIKYAIAYYDKGLPELTENVIHFFMKPFAEFGD-----PYWVSYIRKN 390  
QY 401 VKTFPADFPYSIYETHAASNGSMSEYDKSDGSELSARDLTWVSAAALLTANRNSV 456  
Db 391 MATYADNPLKAAVAFQHPNGMSSEQPSRDGQKQKARDLTWVSISLNAITYREAL 446

RESULT 8

AMYG\_RH10R STANDARD; PRT; 604 AA.  
ID AMYG\_RH10R  
AC P07683;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Glucoamylase precursor (BC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)  
DE (1,4-alpha-D-glucan glucohydrolase).  
OS Rhizopus oryzae (Rhizopus delamar).  
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;  
OC Rhizopus.  
OX NCBI\_TaxID=64495;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SAM0034;  
RA Ashikari T., Nakamura N., Tanaka Y., Kiuchi N., Shibano Y.,  
RA Tanaka T., Amachi T., Yoshizumi H.,  
RT "Rhizopus raw-starch-degrading glucoamylase: its cloning and  
RT expression in yeast".  
RU Agric. Biol. Chem. 50:957-964(1986).  
RN [2]  
RP HOMOLOGY, AND PREDICTED SECONDARY STRUCTURE.  
RA Tanaka Y., Ashikari T., Nakamura N., Kiuchi N., Shibano Y.,  
RA Amachi T., Yoshizumi H.,  
RT "Comparison of amino acid sequences of three glucoamylases and their  
RT structure-function relationships".  
RU Agric. Biol. Chem. 50:965-969(1986).  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
CC glucose residues successively from non-reducing ends of the chains  
CC with release of beta-D-glucose.  
CC -1- MISCELLANEOUS: Rhizopus glucoamylase exists in multiple forms,  
CC Gluc 1, Gluc 2, and Gluc 3, all of which hydrolyze gelatinized  
CC starch at similar rates, but only the largest one (Gluc 1) is able  
CC to adsorb raw starch.  
CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

Query	Subject	Score	DB	Length	Matches	Conservative	Mismatches	Indels	Gaps
QY	ATLDSWLSNENAVTARTALINNGAGAWVSGADSGIVASPSTNDPDYFYTTTSDGLVL	26.2*	Score 725;	DB 1;	Length 604;	36.8*	Pred. No. 1.6e-40;		
Db	STISSMIRKGGGSGISFAMLRINP-----PGSATGFIAASLSTAGPDVYYATMTRAA	165;	Conservative	71;	Mismatches 170;	Indels 42;	Gaps 10;		
QY	25 ATLDSWLSNENAVTARTALINNGAGAWVSGADSGIVASPSTNDPDYFYTTTSDGLVL	84							
Db	168 STISSMIRKGGGSGISFAMLRINP-----PGSATGFIAASLSTAGPDVYYATMTRAA	222							
QY	85 KTLVDLFR--NGDTSLLSTLSTENYISAQAIYVGISNPSGDISSGAGLGEPKFENVDETAYT	141							
Db	223 NVIVVEYNTTSLGNKTLINLVAKDYVTFSVKTKQSTSTVCN-----CLGEPKFNPDSAGYT	276							
QY	142 GSNRPQPDGGRALRATAMIGG-QWLLDNGYTSATDVLPLVRNDLSYVAQYNNQGYD	200							
Db	277 GAWGRPQPDGGRALRATAMIGG-QWLLDNGYTSATDVLPLVRNDLSYVAQYNNQGYD	336							
QY	201 LMEEVNGSSPFTIANQHRALVEGSAFATAVAGSS--CSMCSQSAPEILCYLQSFMTG	258							
Db	337 LMEEVNGVHFITLVMRKGLIGADFAKRNQDSTASTSYSTASTANKISSFWSSNNW	396							
QY	259 ANPDS-----SRGSKDANTL-----LGSIHFTFDPACDDSTFQPCSPRALNHNKEV	308							
Db	397 IQVGSQVTVGVGSKGLDVTSLTLANLGSV-----DDGFFTGSGEKILATATAVA	447							
QY	309 FRSIYTLNDGSLDSEFAVAVGRYPEDTY-----YNGNPMFLCTLLAAEQLYDALYQ	363							
Db	448 FASLYPIPIKNNPSYLGNSIGRYPEDTYVNGNSQNSQNSFLAVTGYAEALYIRAIKEM	507							
QY	364 SLEWTDVSLDPEFKKLYSDAAG-TYSSSSSTYSSTVDVKTPADQFVSIVETHAASNG	422							
Db	508 GVTYSSISLPEFFKRFDSSTSGKTYVTGSDPNNNIALAADRFLSTVQIHAHNGSL	567							
QY	423 SEQYDKSDGEQLASARDLTMSYALITAN	450							
Db	568 ABEFPRTTGLSTGARDLTMSHSLITAS	595							

```

DE GLUCOAMYLASE GLAI precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
DN glucosidase) (1,4-alpha-D-glucan glycohydrolase).
GN GLAI.
OS Saccharomycopsis fibuligera (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycopsidaceae; Saccharomycopsis.
OX NCBI_TaxId=4944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KZ;
RA MEDLINE=92137640; PubMed=1840532;
RT Hostinova E., Balanova J., Gasperik J.;
RT "The nucleotide sequence of the Glucoamylase gene GLAI from
RL Saccharomycopsis fibuligera KZ."
RL FEWS Microbiol. Lett. 67:103-108(1991).
RN [2]
RP REVISIONS.
RC STRAIN=KZ;
RA Hostinova E.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES. ONLY
CC DIFFERS IN 11 POSITIONS WITH GLUCOAMYLASE GLUI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on iteans
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X58117; CAA41120.1; -.
DR HSSP; P08017; IAYX.
DR InterPro; IPRO00165; Glyco_hydro_15.
DR InterPro; IPRO08928; Glyco_trans_gbp.
DR Pfam; PF00723; Glyco_hydro_15; 1.
DR PRINTS; PR00736; GMYDRSLSE15.
DR PROSITE; PS00820; GLUCOAMYLASE; 1.
KM Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KM Signal.
FT CHAIN 1 27
FT BINDING 28 519
FT ACT_SITE 234 234
FT ACT_SITE 237 237
FT ACT_SITE 238 238
FT CARBOHYD 115 115
FT CARBOHYD 127 127
FT CARBOHYD 205 205
SQ SEQUENCE 519 AA; 57542 MW; A15A009A7640053C CRC64;
Query Match 24.2%; Score 671.5; DB 1; Length 519;
Best Local Similarity 35.2%; Pred. No. 4.3e-37;
Matches 173; Conservative 73; Mismatches 169; Indels 77; Gaps 16;
24 RAITDSWLSNEATYARPAIALINIGADGAWMSGDGIVASPSPTNDPIFYTTWRDGL- 82
Db :::: : : : : : : : : : : : : : : : : : : : : : : : : : :
42 RTDETFLDKQKCVSLVLLNLIYPPGQFDVGVTIASPTSISNPYYQWTRDSAIT 101
Qy ---VLKTLTVLDLFRRGDTSILTSTIENTYISAQAIVOGISNPSG-DLSSGAGLGEPKNVD 137
Db FLTYVLSELSD--NNFNPTLLAKAVETIYINTSYNLORTSNPSSFPDENHKGGLGEKFENTDG 159
Qy TAVTSGMRPGORDGPALRATM-----IGFGWL-----DNGYSTATDIWPLVR 184
Db SAYTGAMGRPNDDPALRAYAISRYLNDVNSLNKGLVLTDSGDISNST-EDIKYNLIK 218
Qy NDLSYAQAQYNNQTGYDLMEENGGSFPTIAVQHALL-----EGSAFATAVGSSCS 235
Db SAYTGAMGRPNDDPALRAYAISRYLNDVNSLNKGLVLTDSGDISNST-EDIKYNLIK 218
Db PDLEEVGVGWGSTPDPMEEENGHGFHTSLVOQALVAADIASKFDGDPANLSTAS 278

```

QY 236 WCDSDAELICVIOSEFTGS-----FILANPD-----SSRSKQDANTLGSHTFPD 281  
 Db 279 -----TLESYSGSDGDFVNTDVNHIYENPDLQONSROGSDSATYIGPLLTTHD 327  
 QY 282 PBAACDSTFQCPSPRALAHKEVDSFRSITYLTINDLSSEAVNARYEPEDRY-----Y 336  
 Db 328 -IGSSSTPPDVNDVEYVQSYLLLEENKORYSVNSAY--SAGAIIRYPEDVYNGDSS 384  
 QY 337 NGNPMFLCTLAALAEQYLDALYQMDKQGLEVT--DVSIDFEFKALYSDAAT----- 384  
 Db 385 EGNPMFLATVAAQVPYKALYVD-AKSASNDITINKINYDFPNKXIYVLSLTINSGYSSDS 443  
 QY 385 GTYSSSSSTYSIVDAVKTPADGFSIVETHAASNGSMSEQYKSDGQSLARDLTWSYA 444  
 Db 444 VTIKSGSDEFNTVADNIVTFEGDSFLQVTLIDHINDGSLNQLNNTGYSTSAVSLTWSG 503  
 QY 445 ALLTANNRRNSV 456  
 Db 504 ALLTAIRLRNKV 515

RESULT 10  
 AMYG\_SACPI STANDARD; PRT; 519 AA.  
 ID P08017;  
 AC 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE glucosylase GluI precursor (EC 3.2.1.3) (Glu can 1,4-alpha-  
 DE glucosidase) (1,4-alpha-D-glucan glucosylase).  
 GN GluI.  
 OS Saccharomycopsis fibuligera (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycopsidaceae; Saccharomycopsis.  
 OX NCBI\_TaxId=4944;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8730799; PubMed=3114236;  
 RA Itoh T., Ohtsuki I., Yamashita I., Fukui S.;  
 RT "Nucleotide sequence of the glucosylase gene GluI in the yeast  
 RT Saccharomycopsis fibuligera".  
 RL J. Bacteriol. 169:4171-4176 (1987).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).  
 RX MEDLINE=98437615; PubMed=9757101;  
 RA Sevcik J., Solovicova A., Hostinova E., Gasperik J., Wilson K.S.,  
 RA Dauter Z.;  
 RT "Structure of glucosylase from Saccharomycopsis fibuligera at 1.7-A  
 RT resolution".  
 RL Acta Crystallogr. D 54:854-866 (1998).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 CC glucose residues successively from non-reducing ends of the chains  
 CC with release of beta-D-glucose.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES. ONLY  
 CC DIFFERS IN 11 POSITIONS WITH GLUCOAMYLASE GLAI.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL: L25641; AAA83997.1; -;  
 DR EMBL: M17355; AAA34649.1; -;  
 DR PDB: 1AYX; 13-MAY-98.  
 DR InterPro: IPR000165; Glyco\_hydro\_15.  
 DR InterPro: IPR008928; Glyco\_tran\_6hp.  
 DR Pfam: PF00723; Glyco\_hydro\_15; 1.  
 DR PRINTS: PR00736; GLHYDRLASE15.  
 DR PROSITE: PS00820; GLUCOAMYLASE; 1.

KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
 KW signal; 3d-structure.  
 FT SIGNAL 1 27  
 FT CHAIN 28 519  
 FT BINDING 166 166  
 FT ACT SITE 234 234  
 FT ACT SITE 237 237  
 FT ACT SITE 238 238  
 FT CARBOHYD 115 115  
 FT CARBOHYD 127 127  
 FT CARBOHYD 127 127  
 FT CARBOHYD 205 205  
 FT TURN 30 31  
 FT TURN 36 37  
 FT TURN 42 43  
 FT HELIX 45 61  
 FT TURN 62 63  
 FT STRAND 64 65  
 FT TURN 67 68  
 FT TURN 71 71  
 FT TURN 76 77  
 FT STRAND 79 79  
 FT STRAND 85 85  
 FT STRAND 89 89  
 FT STRAND 93 94  
 FT HELIX 95 111  
 FT TURN 112 113  
 FT HELIX 116 133  
 FT TURN 134 134  
 FT STRAND 137 137  
 FT TURN 138 139  
 FT STRAND 140 140  
 FT TURN 142 143  
 FT TURN 145 151  
 FT STRAND 154 155  
 FT TURN 157 158  
 FT STRAND 161 161  
 FT TURN 171 172  
 FT HELIX 173 191  
 FT TURN 193 194  
 FT TURN 199 200  
 FT TURN 202 203  
 FT HELIX 209 215  
 FT TURN 216 216  
 FT HELIX 217 227  
 FT TURN 228 229  
 FT STRAND 232 233  
 FT TURN 235 236  
 FT STRAND 241 241  
 FT HELIX 244 263  
 FT TURN 264 265  
 FT HELIX 267 285  
 FT TURN 287 289  
 FT STRAND 292 293  
 FT TURN 294 297  
 FT STRAND 298 299  
 FT HELIX 303 307  
 FT TURN 308 309  
 FT TURN 316 316  
 FT HELIX 317 325  
 FT TURN 328 329  
 FT TURN 338 339  
 FT HELIX 341 357  
 FT HELIX 359 363  
 FT TURN 374 375  
 FT STRAND 378 378  
 FT STRAND 383 383  
 FT STRAND 389 389  
 FT STRAND 390 410  
 FT TURN 411 411  
 FT STRAND 414 417  
 FT HELIX 418 420  
 FT HELIX 421 427  
 FT TURN 428 428

FT HELIX 431 433  
 FT TURN 434 434  
 FT TURN 436 437  
 FT HELIX 438 440  
 FT STRAND 443 446  
 FT TURN 448 449  
 FT HELIX 451 474  
 FT TURN 477 478  
 FT STRAND 484 485  
 FT TURN 489 492  
 FT STRAND 492 493  
 FT TURN 496 497  
 FT HELIX 499 519  
 SQ SEQUENCE 519 AA; 57539 MW; BE73035AD1B7652 CRC64;

Query Match 24.0%; Score 664.5; DB 1; Length 519;  
 Best Local Similarity 34.8%; Pred. No. 1,3e-36;  
 Matches 171; Conservative 73; Mismatches 171; Indels 77; Gaps 16;

24 RATLWSLSEATVARTALINNIAGDAGWVGADSGIVVSPSTNDPDTYFYTWRDGL-82  
 42 RTDLFFFLDKQKESLYLLQNTAYPBGQFNNGVPGVIVASPTSPNDYTYOTRDSAIT-101  
 83 ---VLKTLVDLFRNGDTSLLSTIENTYISQAIVQGI5NPSG--DLSSAGIGEPKENVDE-137  
 102 FLTVLSELD--NNFNTTLAKAVEYINTSYNLTQRTSPSGSPFDEHMKGLGEPKENVTDG-159  
 138 TATYSGMRQRQGRPALRATM-----IGFGML-----DNGYTSRTATDIWPLVR-184  
 160 SATYTGAMGRPQNDGPALRAVAISRVLNDVNSLEGKLVITDSGDINFSST-EDLYKNIIK-218  
 185 NDLSYVAQVYNOTGYDLMEVNGSSPFTIIVORHALV-----EGSAFATAVAGSSCS-235  
 219 PDLEVIYIGYDSTGCFDLMENQGRHFTSLVQOKALAVANDIAKSPDGDPAVTLSTAS-278  
 236 WCD5QAPELICYLQSEFWTGS-----FILANFD-----SSRSGKANTLLGSIHTPD-281  
 279 -----TLESYLSGSDGCFVNTDVNHIYENPDLQONSQGLDSATYIGPLLTTHD-327  
 282 PEACACDSTQPSPALANHKEVDSFRSITLNDGLSSEAVAVRPEPTY-----Y-336  
 328 -IGESSSTPDPVNEVYLQSYLLLEBNKDRYSVNSAY--SAGAALGRYEDVDYNGDGS-384  
 337 NGNPFICTAAAEOLYDALYQMDKQGLEVT--DVSLDF-----KALYSDAT-384  
 385 EGNPFLATAYAAQVRYKAYD-AKASANDITINKIYDFNKIYDLSITNSAYQSSDS-443  
 385 GYVSSSSSTYSIVDAVKTFAFGFVSIVETHAASNGSMSEQYKSDGOLSAARDLTW5YA-444  
 444 VTIKSGSDENVTYADNLVTFGDSFLQYILDHINDGSLNQLNRKYGYSTGCAVSLTWSSG-503  
 445 ALLTANNRRNSV-456  
 504 ALLEAIRLRNKV-515

RESULT 11  
 AMYG\_ARXAD STANDARD; PRT; 624 AA.  
 ID AMYG\_ARXAD STANDARD; PRT; 624 AA.  
 AC P42042;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Glucanase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)  
 DE (1,4-alpha-D-glucan glucanhydrolase).  
 GN GAA.  
 OS Arxula adeninivorans (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Dipodascaceae; mitosporic Dipodascaceae; Arxula.  
 NC NCBI\_TaxID=37620;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LS3;

RA Bui Minh D., Kunze I., Foerster S., Wartmann T., Horstmann C.,  
 RA Manteuffel R., Kunze G.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 CC glucose residues successively from non-reducing ends of the chains  
 CC with release of beta-D-glucose.  
 CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL, curation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; Z46901; CAA66997.1; --  
 DR HSSP; P04064; IGAI.  
 DR InterPro; IPR005036; CBM\_21.  
 DR InterPro; IPR000165; Glyco\_hydro\_15.  
 DR InterPro; IPR008928; Glyco\_crbm\_6hp.  
 DR Pfam; PF00723; Glyco\_hydro\_15; 1.  
 DR PRINTS; PR00736; GLHYDRLASE15.  
 DR PROSITE; PS00820; GLUCOMYLA5; 1.  
 KM Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
 KW signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 624  
 FT ACT SITE 340 340  
 FT ACT SITE 343 343  
 FT ACT SITE 344 344  
 FT CARBOHYD 54 54  
 FT CARBOHYD 70 70  
 FT CARBOHYD 98 98  
 FT CARBOHYD 111 111  
 FT CARBOHYD 168 168  
 FT CARBOHYD 267 267  
 FT CARBOHYD 333 333  
 FT CARBOHYD 460 460  
 FT CARBOHYD 582 582  
 SQ SEQUENCE 624 AA; 69890 MW; 74C2CDBF843FF71B4 CRC64;

Query Match 23.2%; Score 641.5; DB 1; Length 624;  
 Best Local Similarity 32.8%; Pred. No. 5.3e-35;  
 Matches 154; Conservative 90; Mismatches 172; Indels 53; Gaps 13;

27 LQSWL--SNEAT--VARTALINNIAGDAGWVGADSGIVVSPSTNDPDTYFYTWRDGL-82  
 160 ITWMLKPSNDQSGIAKSLFNNIN-----IPGAAGTVIAAQSSEPDYATVWVDASL-214  
 83 VLKTLVDLFRNGDTSLLSTIENTYISQAIVQGI5NPSGDLSSAGIGEPKENVDETAAYTG-142  
 215 VMDVVRILYSSAKSEKRLYEKILFYAKAG-AOEONDPALISGGEPEFYLNNTAFITG-273  
 143 SWGRPQNDGPALRATMIGFGWLING-----YTSTATDIWPLVRNDSLVAQ-192  
 274 SWGRPQNDGPATATILIEANAYLANGSGDQTVREQLDSDKYPQAP-IRKOLQPVAS-332  
 193 YNQGYDLMEEVNGSSPFTIIVORHALVGSFAFATVAGSS--CSWCD5QAPELICYLOS-250  
 333 NMSSPFDLMBEBSAHFYTRLVQKALLGADFANDMGHELSDLTKTQASLSPTLPE-392  
 251 FWTGSLTANFD-----SSRSGKANTLLGSIHTDPPEACDSTQPSPALANHKE-304  
 393 FMD5ARQLILYEGPVLGRGYKIDISVVLGWAGY-----ANDNVESTYNDIILTAAYQ-447  
 305 VVDSFISITLNDGLSDS---AVAVGRYPEPTY-----YNGNPFICTAAAEOLYDA-355  
 448 VSTSFLDVTKVANTTSDBSGKPLGIVGRYPEPTYVGVGTSGQNPVYLLTMAAEFLYS-507  
 356 LYQMDKQGLEVTVDLDFPKALYSDA---ATGYSSSSSTYSIVDAVKTFAFGFVSIV-412







RA Shima H., Inui M., Akada R., Yamashita I.;  
 RT "upstream regions of the yeast glucosylase gene which are required  
 for efficient transcription.";  
 RL Agric. Biol. Chem. 53:749-755(1989).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 glucose residues successively from non-reducing ends of the chains  
 with release of beta-D-glucose.  
 CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X02649; CAA26487.1; ALT\_INIT.  
 DR EMBL; D00428; BAA00332.1; -.  
 DR PIR; A21896; ALBYG.  
 DR HSSP; P08017; IAYX.  
 DR InterPro: IPR000165; Glyco\_hydro\_15.  
 DR InterPro: IPR008928; Glyco\_trans\_6np.  
 DR Pfam; PF00723; Glyco\_hydro\_15; 1.  
 DR PRINTS; PR00736; GLHYDRLASE15.  
 DR PROSITE; PS00820; GLUCOMYLAZE; 1.  
 KM Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
 KW Signal; Multigene family.  
 FT SIGNAL 1 21  
 FT CHAIN 22 767 GLUCOMYLAZE S1.  
 FT DOMAIN 22 347 SER/THR-RICH.  
 FT DOMAIN 348 691 H SUBUNIT.  
 FT BINDING 692 767 Y SUBUNIT.  
 FT ACT\_SITE 518 518 CATALYTIC BASE (BY SIMILARITY).  
 FT ACT\_SITE 521 521 GENERAL ACID CATALYST (BY SIMILARITY).  
 FT ACT\_SITE 522 522 INTERACT WITH SUBSTRATES (BY SIMILARITY).  
 FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 546 546 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 650 650 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 741 741 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 767 AA; 82488 MW; A5F29E2427EDB593 CRC64;  
 Query Match 18.8%; Score 522; DB 1; Length 767;  
 Best Local Similarity 30.5%; Pred. No. 5,5e-27;  
 Matches 139; Conservative 80; Mismatches 177; Indels 60; Gaps 15;  
 QY 1 MSFRSLALSGVCTGLANVYSKATLDSWLSMEATYARAILNNGADGAWYSGADSGI 60  
 DB 313 VQLRDVLAMGTYVYD-SNGAMDSALAEWLRQKVKYSIRIENIGPSAVYPS-ILPGV 370  
 QY 61 VVASPTDNDYFYTYTRDSGLVLTVDLFRNGDTSLSTIENYISAQIVOGISNPSG 120  
 DB 371 VVASPGQTHDYFYQWIRDSALITNSIVS--HSADPA-ITTLQYLVNVSFHLRTNN--- 424  
 QY 121 DLSSGAG-----LGEKRVDEIATYTSWGRPPRDGRLATAMIGFQWLLDNG--- 170  
 DB 425 TLGAGGYTNDVALGDPKKNVNTAFTEPWRGPONDGRLAIRAIKIIDYIKOSGTDL 484  
 QY 171 -----TSTATDVLVPLVRNDLSVAQYMNQGTVDLMEEVNGSSFFTIAYQHAALVSGSA 225  
 DB 485 GAKYPPQSTH-DLFPDDLVKRDLFITIDHWNSSGDLMEEVNGHFFLLVQLSALVDRSL 543  
 QY 226 FATAVSSGSCWCD---SQAPEILCYL-----QSFWTGSFIIANF-----DSSRSGKDA 270

DB 544 YFNASERSPPFYBELRQTRRDISKFLVPDANGFINKY---NIYETPMIADTLRSGLDI 600  
 QY 271 NTLLGSIHTFDEPAACDDSTFPQCSFPLANKKEVDSFRSITYTLNDGUSDSEAVAVGRY 330  
 DB 601 STLLANVTYHADPSA-SHPFINDPVAIVNTLHMLHNRISITYPINDSKMTGIALGRY 659  
 QY 331 PEDTY-----YNGNPWFCTLAAEQLYALYQMDKQSLBYTVDSLDFPKALYSDAATG 385  
 DB 660 PEDVYDYGVGGEENPVALTCAASTLYQLYRHISEQHDLPVPMNNDGSAFNMFSELVRS 719  
 QY 386 TYSS-----SSSTYSIVDAVTFPFDGAV 409  
 DB 720 NTLLGNDGUYLLEFNTPAFQTLQKIFQLADSF 755  
 RESULT 14  
 ID AMY1\_SACDI STANDARD, PRT, 768 AA.  
 AC P29760;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Glucosylase S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)  
 DE (1,4-alpha-D-glucan glucosylase) (Gall).  
 GN SYA2 OR DEX1.  
 OS Saccharomyces diastaticus (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_Taxid=41870;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94288586; PubMed=8017901;  
 RA Kim K., Bajszar G., Lee S.Y., Knudsen F., Mattoon J.R.;  
 RT "Cloning of a new allelic variant of a Saccharomyces diastaticus  
 glucosylase gene and its introduction into industrial yeasts.";  
 RL Appl. Biochem. Biotechnol. 44:161-185(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91276266; PubMed=2055484;  
 RA Lambrechts M.G., Pretorius I.S., Solliet P., Marmur J.;  
 RT "Primary structure and regulation of a glucosylase-encoding gene  
 (STA2) in Saccharomyces diastaticus.";  
 RL Gene 100:95-103(1991).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 glucose residues successively from non-reducing ends of the chains  
 with release of beta-D-glucose.  
 CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M60650; AAA5107.1; -.  
 DR EMBL; M90490; AAA20560.1; -.  
 DR PIR; J00474; J00474.  
 DR HSSP; P08017; IAYX.  
 DR InterPro: IPR000165; Glyco\_hydro\_15.  
 DR InterPro: IPR008928; Glyco\_trans\_6np.  
 DR Pfam; PF00723; Glyco\_hydro\_15; 1.  
 DR PRINTS; PR00736; GLHYDRLASE15.  
 DR PROSITE; PS00820; GLUCOMYLAZE; 1.  
 KM Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
 KW Signal; Multigene family.  
 FT SIGNAL 1 21  
 FT CHAIN 22 768 GLUCOMYLAZE S2.  
 FT DOMAIN 22 348 SER/THR-RICH.  
 FT DOMAIN 349 692 H SUBUNIT.  
 FT BINDING 693 768 Y SUBUNIT.  
 FT BINDING 456 456 SUBSTRATE (BY SIMILARITY).

FT ACT SITE 519 519 CATALYTIC BASE (BY SIMILARITY).  
 FT ACT SITE 522 522 GENERAL ACID CATALYST (BY SIMILARITY).  
 FT ACT SITE 523 523 INTERACT WITH SUBSTRATES (BY SIMILARITY).  
 FT CARBOHYD 35 35 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 435 435 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 514 514 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 721 721 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 742 742 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 164 164 MISSING (IN REF. 1).  
 FT CONFLICT 624 624 D -> N (IN REF. 1).  
 SQ SEQUENCE 768 AA; 82586 MW; 3PAC172C128A0CF CRC64;

Query Match 18.8%; Score 521; DB 1; Length 768;  
 Best Local Similarity 30.5%; Pred. No. 6,4e-27;  
 Matches 139; Conservative 80; Mismatches 177; Indels 60; Gaps 15;

QY 1 MSRRSLALSGVCTGLANYSKRATLDSWLSNEATYARTRAILNLITAGDAMVSGADSGI 60  
 DB 314 VOLRDVLANGVVYD-SNGAMDSPLLEWLQROKKVIRIFENIGPSAVYPS-ILPGV 371  
 QY 61 VVAPSTDPNDPYFTWTRDGLVLTVDLFRNGDTSLSTIENYISAQAIIVOGISNPSG 120  
 DB 372 VIAPSPQTHDYPQWIRDSALTNISVS--HSDPA-ITELQYLVANSHLQRTNN--- 425  
 QY 121 DLSSGAG-----LGEPRNVDFTAYTSGWRQDGPALRATAMIGFQWLLDNG--- 170  
 DB 426 TLGAGIYTDVVALGDPKNVNDTAFTEPWRQDGPALRSIALIKIIDIYIKOSGTDL 485  
 QY 171 -----YTSTATDIWPLVRNDLSVAQYMNQYGVDLMEVNGSSFFLIANQHRALVGS 225  
 DB 486 GAKYPFQSTA-DIFDDIVRMDLRFIDHMNSGDLMEVNGHFFLIVQLSAVDRSL 544  
 QY 226 FATVAGSSGCMCD--SQAPEILCYL---QSFWTGSFIILANF-----DSSRSGKDA 270  
 DB 545 YFNABSESSPVEELRQTRDISKEFLVDPANGFNGKY--NYVETPMADTIRSGLDI 601  
 QY 271 NTLLGSIHTFDPEAACDSTFQPCSPRALANKEVDSFRSIYTLNDGLSDSEAVAVGRY 330  
 DB 602 STILLAANTVDAPSA-SHLPEPIDDPVAVLNTLHLMHMSIYPIINDSSKNATGIALGRY 660  
 QY 331 PBDTY-----YNGNPMFLCTLAABQYDALYQMDKGSLEYVDVSLDFEALYSDAATG 385  
 DB 661 PEDYDGYGVGEGNPMVLATCAASTLYOLYRHSOCHDLVVPNMNDCSNAPWSLWFS 720  
 QY 386 TYSS-----SSSTYSSIVDAVKTFADGFV 409  
 DB 721 NLTLLGNDGGLILIEFNTPAFNQTIQKIFQLADSF 756

RESULT 15  
 YG10 METJA STANDARD; PRT: 615 AA.  
 ID YG10 METJA  
 AC Q59005;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical glycosyl hydrolase MJ1610 (EC 3.2.1.-).  
 GN MJ1610.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_Taxid=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kierulff A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Gilek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,  
 RA Urrbach T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Coton M.D., Roberts K.M., Hure M.A., Kane B.P., Bordovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii";  
 RL Science 273:1058-1073(1996).  
 CC -1- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U67601; AAB99630.1; -.  
 CC PIR; A64501; A64501.  
 DR TIGR; MJ1610.  
 DR Interpro; IPR000165; Glyco\_hydro\_15.  
 DR Interpro; IPR008928; Glyco\_trans\_6np.  
 DR Interpro; IPR006465; Oligosac\_amy1.  
 DR Pfam; PF00723; Glyco\_hydro\_15; 1.  
 DR TIGRFams; TIGR01577; oligosac\_amy1; 1.  
 DR PROSITE; PS00820; GLUCOAMYLASE; 1.  
 KW Hypothetical protein; Hydrolase; Glycosidase; Complete proteome.  
 FT ACT SITE 406 406 GENERAL ACID CATALYST (BY SIMILARITY).  
 FT ACT SITE 407 407 INTERACT WITH SUBSTRATES (BY SIMILARITY).  
 SQ SEQUENCE 615 AA; 72007 MW; 2B37EB89F0357BES CRC64;

Query Match 8.0%; Score 221.5; DB 1; Length 615;  
 Best Local Similarity 23.1%; Pred. No. 2.9e-07;  
 Matches 101; Conservative 46; Mismatches 125; Indels 165; Gaps 20;

QY 59 GIIVAPSTDPNDPYFTWTRDGLVLTVDLFRNGDTSLSTIENYISAQAIIVOGISNP 118  
 DB 284 GIIVAPSL-HPRYRYWGRD-GSYISALDLF-----GIRNPDPFERMSKIQ----- 331  
 QY 119 SGLSSGAGLGEPRNVDFTAYTGSW-----GRQDGPALRATAM-----IGFQWLL 167  
 DB 332 -----NAD-----GSLQYVYVNGK-----RLTAIQTDQIGSILMMAM 364  
 QY 168 DNGYTAIDIWPLVRNDLSVAQYMNQYGVDLMEVNGSSFFLIANQHRALVGS 211  
 DB 365 DVHYRLT-----GDRKFEYRYWNTIEKANYLRVALNFTPCFDLMEERFGVFAY 414  
 QY 212 TLIVOHRAVESGAFATAV-----GSSCSMCDQAPEILCYLSQFWTGSFIILANFP 262  
 DB 415 TWGATYAGLKCAVSKAVKRDYKDWGKTIIFLGHVPRK-YLE-----D 461  
 QY 263 SSRSKGDANTLLGSIHTFDPEAACDSTFQPCSPRALANKEVDSFRSIYTLNDGLSDS 322  
 DB 462 EERFAKSIINPL-----DKTIDTSILGLSYPNLID-----VDERMIKT-----A 501  
 QY 323 EAV-----AVGRPEPTYNGNPMFLCTLAAB-----EQYDALYQMDKGSLEYVDV 370  
 DB 502 EALEKAFKYGKVGIGRPEDYFGNPMIITTLMLLYRRLKVKKEKDQNA---DI 557  
 QY 371 SLDFPALYSDAATGYSSSSSYSSIVDAVKTFADGFVSVIETHAASNGMSGOYDKSD 430  
 DB 558 YLQSKSLFWMVKYSF-----DGLPFEQIHKE 586  
 QY 431 GEOLASADLTWSYAL 447  
 DB 587 GVPMSAMPLGMSNMF 603

RESULT 16

ID	ICEV_PSEX	STANDARD	PRT	1196 AA.
AC	033479;			
AD	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Ice nucleation protein.			
GN	INAV.			
OS	<i>Pseudomonas syringae</i> .			
OC	<i>Bacteria</i> ; <i>Proteobacteria</i> ; <i>Gammaproteobacteria</i> ; <i>Pseudomonadales</i> ;			
OC	<i>Pseudomonadaceae</i> ; <i>Pseudomonas</i> .			
OX	NCBI_Taxid=317;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=INAV;			
RA	MEDLINE=97462815; PubMed=9323042;			
RA	Schmid D., Pridmore D., Capitani G., Battistutta R., Neeser J.-R.,			
RT	Jann A.;			
RT	"Molecular organisation of the ice nucleation protein Inav from			
RT	<i>Pseudomonas syringae</i> ."			
RL	FEBS Lett. 414:590-594(1997).			
CC	-1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate			
CC	crystallization in supercooled water.			
CC	-1- SUBUNIT: MEMBRANE ENVIRONMENT OR AGGREGATION SEEMS TO BE REQUIRED			
CC	FOR ICE NUCLEATION ACTIVITY.			
CC	-1- SUBCELLULAR LOCATION: Outer membrane (By similarity).			
CC	-1- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS			
CC	OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A			
CC	REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.			
CC	-1- SIMILARITY: Belongs to the bacterial ice nucleation protein			
CC	family.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ).			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; AJ001086; CA04521.1; -			
DR	HSSP; P06620; 11NA			
DR	InterPro; IPR000258; Ice_nucleatn.			
DR	Pfam; PF00818; Ice_nucleation; 61.			
DR	PRINTS; PR00327; ICENUCLEATN.			
DR	PROSITE; PS00314; ICE_NUCLEATION; 42.			
FM	Ice nucleation; Repeat; Outer membrane.			
KT	DOMAIN 172 1147 OCTAPEPTIDE PERIODICITY.			
SO	SEQUENCE 1196 AA; 117991 MW; C9E974CB1731E68 CRG64;			
Query Match	5.8%; Score 159.5; DB 1; Length 1196;			
Best Local Similarity	24.0%; Pred. No. 0.0088;			
Matches 133;	Conservative 60; Mismatches 224; Indels 155; Gaps 29;			
QY	47 GADGAWV-----SGADSGIVVASPSTNDPD-----YFYTWI--RDSGLVLKTLV 88			
DB	357 GADSLIAGYGSTQTSGSDSLTAGYGSTQTAQEGSNLTAGYGSNLTAGYGSSTQTSGLIAGYGS 416			
QY	89 DLFRNGDTSLSTLENIYNSAQAIYQGISNPGDLSG-----AGLGEPKRVNDETA 139			
DB	417 TQTSGSISLTA--GYSSTQTAQEGSNLTAGYGSSTGAYGDSLSIIAGYGSSTQTSGSISA 473			
QY	140 YTGSMGRPQRDGPALRATAMIGFCOMLLDNGVSTPATYDVIWPLVRNDGSYYAQAQYNNQGY 199			
DB	474 LTAIYGSSTQ-----TQAQGSN--LTAIYGSSTGT-----AGADSLIA--GY 510			
QY	200 DLMEEVNGSFFTLAV-OHRLAVEGSAFATVAGSS-----CSWCDSDQA 241			
DB	511 GSTQTSGESLSLTAGYGSSTQTAQEGSNLTAGYGSSTGADSLIAGYGSSTQTSGSDSL 570			
QY	242 PEILCIQSYFTTSGFILANPSSSSKQDANL--LGSIHFPDEPA--ACDDSTFQFGCS 295			
DB	571 TAGYGSSTQTAQGSNLTAGYGSSTQTAQGSNLTAGYGSSTGTAQGSNLTAGYGSST-- 625			

Query	Match	5.7%	Score 159	DB 1	Length 1210
296	PRLANKEVVDV-FRSTIYTLNCLDSSEVAVCRPEDEPYNGNP----	WFLCTLA	AAE	350	
626	--QTAGHKSLLTAGYSGTQTARDQ----	SDLVH-CYSGTGRAGSSSLIAGYSGTQTAS	YK	679	
351	QLYDALQYQMDKQSLLEVTVSLDPFKALYSDPATGTYS	SSSSSTYS	IVDAVKEP	ADFEVS	410
680	SMLTAGY-----	GSTQFAREHSDLVAGYSGTSTAGSSSLIAGYGS	-----	TQTAFK	728
411	IVETHAASNGSMSEQYKSD-----	GEQLSA----	RDLTWSYAL	446	
729	IM--TGYGSTQTQARSDLVAGYSGTSTAGYSSSLIAGYSGTQTAGYSGTTLT	YTGST	785		
447	LTANNRNSVVPASMGETSASVPGTCAATISITYSVTTSPMSIVATS-GTTT	TATP	505		
786	QTA-QENSLITTYGSGTSTAGY-----	SSSLIAGYSGTQTAGYESTL	TLTAGYSGTQTARQ	838	
506	-----TSGSGVTS	-----	KTTAPASKTSTT	528	
839	RSDLVTGYGS-TSTAGYASSLIAGYSGTQTAGYESTLT	875			
RESULT 17					
ICEN_PSEFL	STANDARD;	PRT;	1210	AA.	
AC	P09815;				
DT	01-MAR-1989 (Rel. 10, Created)				
DT	01-MAR-1989 (Rel. 10, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Ice nucleation protein.				
GN	INAM.				
OS	Pseudomonas fluorescens.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;				
OC	Pseudomonadaceae; Pseudomonas.				
OX	NCBI_TaxID=294;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=87040772; PubMed=3774551;				
RA	Warren G.J., Corotico L., Wolper P.;				
RT	"Conserved repeats in diverged ice nucleation structural genes from				
RT	two species of Pseudomonas."				
RL	Nucleic Acids Res. 14:8047-8060(1986).				
CC	-1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate				
CC	crystallization in supercooled water.				
CC	-1- SUBCELLULAR LOCATION: Outer membrane (By similarity).				
CC	-1- DOMAIN: CONTAINS 122 IMPERFECT REPEATS OF THE CONSENSUS				
CC	OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A				
CC	REGIONAL 48-RESIDUE PERIODICITY IS SUGGESTED.				
CC	-1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE				
CC	NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.				
CC	-1- SIMILARITY: Belongs to the bacterial ice nucleation protein				
CC	family.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce">http://www.isb-sib.ch/announce</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL; X04501; CAA28186.1; -.				
DR	PIR; A25547; A25547.				
DR	HSSP; P06620; IINA.				
DR	InterPro; IPR00258; Ice_nucleatn.				
DR	PIfam; PF00818; Ice_nucleation; 61.				
DR	PROSITE; PRO00327; ICENDCLEATN.				
DR	PROSITE; PS00314; ICE_NUCLEATION; 45.				
KW	Ice nucleation; Repeat; Outer membrane.				
FT	DOMAIN 165 1156 OCTAPEPTIDE PERIODICITY.				
SO	SEQUENCE 1210 AA; 118881 MW; 7C4ADDE633F61080 CR664;				

Best Local Similarity 22.2%; Pred. No. 0.0097;  
Matches 128; Conservative 72; Mismatches 246; Indels 130; Gaps 25;  
QY 8 ALSGLVCTGLANYSKRATLDSMLNSNATVARTAILNINIGADGAWSGADSVIVASPT 67  
Db 165 AVYGSITLTC---ANQGLIAGVSTETADSSITLIVAGVSTG--TSGSSSIIAGVST 218  
QY 68 DNDPYFTWRDGLVLTIVDLFRNGDTSLSTENYISAQAVIGISNPSGDLSSG-- 125  
Db 219 G-----TAGSDSLIAGVSTGTAGDSSSLTA---GYGSTQTQAVQNSNLTAGVSTGT 269  
QY 126 -----AGGPKFNVDERTAYGSMGRPQDGLALATMIGRGWLLDNGTSTPTDI 178  
Db 270 GPDSLLIAGVSTGTAGGESSLTAGVSTG-----TQAVGSD--LTAGVSTGT-- 316  
QY 179 VMLVLRNDLSYVAVQYQMTGYDLMEEVNGSSFTIYAVQHRALVGAFAFATVSSCSMCD 238  
Db 317 -----AGSDSSSLIA-----GYGSTQTAGDSSSLTAGVSTQTQAVQNSNLTAGVSTGTAG 366  
QY 239 SQAPEILCYQSFMTG--SFILANFDSRS--GKDNATLLGSIHTPPEACDDSTFOP 293  
Db 367 PDSSLLIAGVSTGTAGGESSLTAGVSTGTQAVQNSDLTAGVSTGT---AGSDSLIAG 422  
QY 294 C-SPRALANKEVVDSTRTYTL-----NDLSSEAVAVGRYEDTYNGNPNW 341  
Db 423 YGSTQTAGGESSLTAGVSTGTQAVQNSDLTAGVSTGTAGDSSSLIAGVSTGTAGD-- 480  
QY 342 FLCTLAALQYDALYQMDKQSLLEVVDVLDLFFKALYDAATGYSSSSSTYSIYDAV 401  
Db 481 -SSLTG-----YGSTQTQAVQNSDLTAGVSTGTAGDSSSLIAGVGS----- 521  
QY 402 KTEPADGFVSIETHAASNGSMSEQYKSDGDEQASARDLTWSYALLTANNRNSVVPASW 461  
Db 522 -----TQT-AGDSSSLTAGVSTGTQAVQNS-DLTAGVSTGTAGS--DSSLIAGY 567  
QY 462 GETSASSVPTCATSAIGT-----YSSV-TYTSWSIYATGTTTATPTG 507  
Db 568 GSTQTQTA--GDSSLTAGVSTGTQAVQNSNLTAGVSTGTAGDSSSLIAGVSTGTQTA--GG 623  
QY 508 SGSVTSTKTTATASK-----TSTTRSGMSL 534  
Db 624 DSSLITAGVSTGTQTAGHGSILTAGVSTGTQAGSSSL 659  
RESULT 18  
YFGG SCHPO STANDARD; PRT; 670 AA.  
ID YFGG SCHPO STANDARD; PRT; 670 AA.  
AC 013854; 042840;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical serine/threonine-rich protein Cl9612.16c in chromosome I precursor.  
GN SPAC19612.16c OR SPAC23A1.01c.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetes; Schizosaccharomycetaceae;  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Scouras J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Haldago J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volkhardt G., Aert R., Robben J., Grymoprez B.,  
RA Welter J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fitz C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Medler H., Wandut R., Farnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Kottler S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armerstrong J., Forbury S.L.,  
RA Gernutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,  
RA Shiparovsky G.V., Usery D., Barrell B.G., Nurse P.,  
RT "The genome sequence of Schizosaccharomyces pombe."  
RL Nature 415:871-880(2002).  
CC - SUBCELLULAR LOCATION: Secreted (Potential).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z97209; CAB10128.1; -  
DR EMBL; AL021813; CAB16975.1; -  
DR GeneDB Spombe; SPAC19612.16c; -  
KM Hypothetical protein; Glycoprotein; Signal.  
FT SIGNAL 1 19  
FT CHAIN 20 670  
FT FT  
FT DOMAIN 120 653  
FT CARBOHYD 77 77  
FT CARBOHYD 159 159  
FT CARBOHYD 204 204  
FT CARBOHYD 224 224  
FT CARBOHYD 274 274  
FT CARBOHYD 297 297  
FT CARBOHYD 327 327  
FT CARBOHYD 351 351  
FT CARBOHYD 370 370  
FT CARBOHYD 381 381  
FT CARBOHYD 405 405  
FT CARBOHYD 424 424  
FT CARBOHYD 435 435  
FT CARBOHYD 459 459  
FT CARBOHYD 478 478  
FT CARBOHYD 489 489  
FT CARBOHYD 513 513  
SQ SEQUENCE 670 AA; 68262 MW; 68E8D1CFB855FA8 CRC64;  
Query Match 5.7%; Score 158; DB 1; Length 670;  
Best Local Similarity 19.7%; Pred. No. 0.0051;  
Matches 119; Conservative 79; Mismatches 245; Indels 162; Gaps 21;  
QY 7 LALSGL-----VCTGL-ANVISKRALTDS-----WLSNEATVAR 39  
Db 4 LITISGLISLAKCAGMEINVPSSSDVWTSGHIEPLEMSVSTDPLOANVWLINVEYRP 63  
QY 40 TA-----ILNNGADGAWSGADSVIVASPTD-NDPYFY-----TWTRDGLVLTIVDL 90  
Db 64 TSTRYIMTINTDFENATF-----PALDLSFGYQISFTSIRDSDVIYAQSGTF 111  
QY 91 FRNGDTSLSLTINYSIAQAVIGISNPSGDLSSGAGLGEPRNV--DETAAYGSMGRPQ 148  
Db 112 YIVGSGISSTTSTQSMSTFTTSQTSNGHSASTISBTAITVANSTYS----- 165  
QY 149 RDGPALATAMIGFGWLLDNGYSTATDIPWPLVLRNDLSYVAVQYNGQGYDLMEVNGS 208  
Db 166 -----ATSSPYSTDV--SVSTGTSTDIV-----TLPPASST 196

QY 209 SFTTIAVHRAALVEGSAFATAVAGSGCWCDSQAPILLCYLQSFMTGSLIANFDSRSRG 268  
 DB 197 SSSSTITNMSIPSSSSFTTTTGS-----PYYNTSGFLPSSVSSASLS 240  
 QY 269 DANTLIGSIHTFPEACDSTQPCSPRALANHKEVDSFNSIYTLNDGLDSEAVAVG 328  
 DB 241 SSSVLEPSTSIIT-----STSTPVTWSSSS-----LSSFTPSYSTNLTTTGSTTTT-- 284  
 QY 329 RYEDDYNYGNPMFLCTLAAAEQLYDALYQMDQKQSLFETVDLDFPKLLYSNAAAGT-- 386  
 DB 285 ---GSATVSSSPFPYSSSVIPTSVPSSVSFTSSSSSYTTLLTASNTSVTTTGTTGSAT 341  
 QY 387 -----YSSSSSTYSIYDAVATPADGFVSIETHAASNGSSEQ----- 425  
 DB 342 FTSSPPFYSSSVIPTSVPSSVSFTSSNSYTTLLTASNTSITTYGTGATFTSSPP 401  
 QY 426 -YDKSD-----GEQLSARDLTWSYAAALTANNRRNSVVPASWGETSASSVP-----GT 472  
 DB 402 FYNSNSVVIPTSVPSSVSFTSSNSYTTLLTASNTTFTGTGATFTSSPPYSSNS 461  
 QY 473 CATSAIGTYSSVTV--TSMPSIVATGCTTTATPTGSGSVTSTKTATASKT-----ST 526  
 DB 462 VIPTSA PSSVSFTSSNSYTTLLTASNTTFTGTGATFTSSPPYSSNSIIVPTT 521  
 QY 527 TTRSG 531  
 DB 522 VRTSG 526

## RESULT 19

YS89 CAEEL STANDARD; PRT; 3178 AA.

AC 009624; 009625; 0969D4;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical protein ZK945.9 in chromosome II.  
 GN ZK945.9/ZK945.10.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Briscot N2;  
 RA Wilkinson-Sproat J.;  
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP REVISIONS.  
 RA Durbin R.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 CC 1- SUBCELLULAR LOCATION: Integral membrane protein (potential).  
 CC 1- SIMILARITY: Contains 1 GPS domain.  
 CC 1- SIMILARITY: Contains 1 PLAT domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; Z48544; CAB70192.1; -;  
 DR EMBL; Z48582; CAB70192.1; JOINED.  
 DR EMBL; Z48582; CAB70201.1; -;  
 DR EMBL; Z48544; CAB70201.1; JOINED.  
 DR WormRep; ZK945.9; CE25697  
 DR InterPro; IPR002111; Cat\_channel\_TripL.  
 DR InterPro; IPR005821; Ion\_trans.  
 DR InterPro; IPR001024; Lipoxxygenase\_LH2.  
 DR InterPro; IPR003915; PKD 2.  
 DR InterPro; IPR000203; PKD\_cys\_rich.

DR InterPro; IPR008976; PLAT\_LH2.  
 DR Pfam; PF01825; GPS\_1.  
 DR Pfam; PF00520; ion\_trans\_1.  
 DR Pfam; PF01477; PLAT\_1.  
 DR PRINTS; PR01433; POLYCYSTIN2.  
 DR SMART; SM00303; GPS\_1.  
 DR SMART; SM00308; LH2; 1.  
 DR PROSITE; PS50095; PLAT; 1.  
 KW Hypothetical protein: Transmembrane.  
 FT TRANSMEM 13  
 FT TRANSMEM 30  
 FT TRANSMEM 51  
 FT TRANSMEM 73  
 FT TRANSMEM 2139  
 FT TRANSMEM 2161  
 FT TRANSMEM 2348  
 FT TRANSMEM 2367  
 FT TRANSMEM 2390  
 FT TRANSMEM 2412  
 FT TRANSMEM 2451  
 FT TRANSMEM 2468  
 FT TRANSMEM 2483  
 FT TRANSMEM 2505  
 FT TRANSMEM 2567  
 FT TRANSMEM 2589  
 FT TRANSMEM 2836  
 FT TRANSMEM 2858  
 FT TRANSMEM 2939  
 FT TRANSMEM 2961  
 FT TRANSMEM 2976  
 FT TRANSMEM 2998  
 FT TRANSMEM 3038  
 FT TRANSMEM 3060  
 FT TRANSMEM 266  
 FT TRANSMEM 1196  
 FT TRANSMEM 1105  
 FT TRANSMEM 1241  
 FT TRANSMEM 2071  
 FT TRANSMEM 2120  
 FT TRANSMEM 2182  
 FT TRANSMEM 2305  
 FT TRANSMEM 3178 AA; 344726 MM; F8239436D0366CD CRC64;

Query Match 5.7%; Score 158; DB 1; Length 3178;  
 Best local Similarity 20.1%; Pred. No. 0.041;  
 Matches 108; Conservative 81; Mismatches 215; Indels 134; Gaps 19;

QY 32 SNEATVARTAIUNNICADAWVSGADSVVASPSPT-----DNPQVFTYTRSGVLVKT 86  
 DB 508 STPSSSTASSSVSTASSTQSTSTQSSSTTYKSETTSSDGNPDYFEKATTTTYDST 567  
 QY 87 LVDLFRNGDSTLS-----TIENISAQAIVQGISNPSGDLSSGAGLGPKNVD 136  
 DB 568 SSVNLINSLGLIIGVQTSIECTSPSSNVST-----TKD 602  
 QY 137 ETATYSGWCRPORDEPALATAMIGQWLLDNGYVSTATDVLVLRNDS----- 188  
 DB 603 GACFTKSVMP-RLGGTYPASTFVGPG-----NYTFRAT-----MTDDKYYTYTAVY 650  
 QY 189 YVAQVWNGCYDLMEEVNSSPFTIAVGHRAALVEGSAFATAVAGSGCWCDSQAPILLCYL 248  
 DB 651 YIOEYSSIT-----ISESSTSAVASTSTSTPSTPSSTLS-- 685  
 QY 249 QSEFTGSLIANFDSRSRGKANTLLGSIHTFPEACDST-----FQPCSPRA 298  
 DB 686 TSVTTPSSSTRSDSTTSAGSTTLQESTTISEESTTSDTTSSTSSPSSTTA 745  
 QY 299 LANHKEVDSFNSIYTLNDGLDSEAVAVAGRPEDTYNGNPMFLCTLAAAEQLYDALYO 358  
 DB 746 DSTSTLSVQFD--FILDGSLSWNET-----RHNEDE-----SINIVPLPTAATP 788  
 QY 359 MDKQSLFETVDLDFPKLLYSNAAAGTSSS--STYSIYDAVATPADGFVSIETHA 416  
 DB 789 TKSQTFEGRNVTSTP-LIESTCLNNTVLANVYSSNP-----IQPLEFL 838  
 QY 417 ASNGSSEYQYDSDGQLSARDLTWSYAAALTANNRRNSVVPASWGETSASSVPGTCAAT 476  
 DB 839 VGIGTYEPIINMTDLTMQV-----VSHIFILNAVADS-----TSTSEVSTTSGSSSE 889  
 QY 477 SAIGYSSVTVTSMPSIVATGCTTTATPTGSGSVTSTKTATASKT-STTRSGMS 533  
 DB 890 SAISTSGIESTSTLEASTTDAQSDSTSTSDGTTSDTSDSNTSTSTSDSGLS 947

## RESULT 20

ID ICEN ERWHE STANDARD; PRT; 1258 AA.  
 AC P16239;  
 DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE ice nucleation protein.  
 GN ICEB.  
 OS Erwinia herbicola.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Pantoea.  
 OX NCBI\_TaxId=549;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=M1;  
 RX MEDLINE=90152370; PubMed=2515997;  
 RA Warren G.J., Corotto L.V.;  
 RT "The consensus sequence of ice nucleation proteins from Erwinia  
 herbicola, Pseudomonas fluorescens and Pseudomonas syringae.";  
 RL Gene 85:239-242(1989).  
 CC -1- FUNCTION: ice nucleation proteins enable bacteria to nucleate  
 CC crystallization in supercooled water.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -1- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE  
 CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE  
 CC PERIODICITY IS SUPERIMPOSED.  
 CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE  
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.  
 CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein  
 CC family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M26382; AAA24823.1; -.  
 DR PIR; J00188; J00188.  
 DR HSSP; P06620; IINA.  
 DR InterPro; IPR000258; Ice\_nucleatn.  
 DR Pfam; PF00818; Ice\_nucleatn. 65.  
 DR PRINTS; PR00327; ICENUCLEATN.  
 DR PROSITE; PS00314; ICE\_NUCLEATION. 45.  
 KM Ice nucleation; Repeat; Outer membrane.  
 FT DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.  
 SQ SEQUENCE 1258 AA; 125084 MW; 590E8A130077BBD4 CRC64;  
 Query March 5.7%; Score 157; DB 1; Length 1258;  
 Best Local Similarity 22.1%; Pred. No. 0.014;  
 Matches 134; Conservative 63; Mismatches 239; Indels 162; Gaps 25;  
 QY 47 GADGAWVGADSGIVASPTDNDPYWTRDGLVLTVDLFRN-----GDTSL 99  
 DB 299 GADSSLLAGVSTGTAGEESTGTGTAGVSTGTAKG-----SDLTAGVSTGTAGDSS-- 350  
 QY 100 STLENYISADIVOGISNPSG-----DLSSG-----AGLGEPEKN 134  
 DB 351 SLIAGVSTGTAGDSSLTAGVSTGTAKGSDLTAGVSTGTAGDSSLIAGVSTGT 410  
 QY 135 VDTEAVYSGMRPORDGPALATAMIGRCGLWLNNGTSTATDVLWPLVNRDLSYVAQW 194  
 DB 411 GEESTGTAGVSTGT-----TAKGSD-----LTAGVSTGT-----AGDSSLI-- 450  
 QY 195 NOTGYDLMEEVNGSSFTTIAVQHRALVGSFAFATAVSSCSGCDQAPELICY--LQSF 252  
 DB 451 ---GYSTGTAGDESSLTAGVSTGTAKGSDLTAGVSTGTAGVSTGTAGVSTGTAG 507  
 QY 253 TGSFTILANFSSRSRGKANTLL---GSIHTFDPACADSTFPQCSPPALANHKEVDS- 308  
 DB 508 YGSLTLAGVSTGTAKHESDLITGYGSTT---AGANSSLIAGVSTGTATSYNSVLTAG 563  
 QY 309 FRSLTYLTNDGL-----SDSEAVN-VERVEDPYTNGN-NWPFCLTAAARD-- 351  
 DB 564 YGSTGTARSGSDLTAGVSTGTAGDSSLTAGVSTGTATSYHSSLTAGVSTGTARQSV 623

QY 352 -----LYDALYQNDKGSLEAVTVSLDFFKALYS 380  
 DB 624 LTTGYSTGTAGADSSLIAGVSTGTAGVNSILTAY-----GSTPAEGSDLTAGVGS 678  
 QY 381 DATGTVSSSSSYSSIVDAVTFADGFVSIVETHAASNGSMEQYDKSDG----- 431  
 DB 679 TSTAGADSSLIAGVGS-----TGTAGVNSILTAYGSGTGTAGGSDLTSGVSTTAGA 732  
 QY 432 -EQLSA---RDLTWSYAALITA-----NNRNSVVPASWGTSASSVPTCAATSAIGT 481  
 DB 733 DSSLIAGVSTGTATSHSSLTAGVSTGTARQESVLTGTGSGTSTAG-----ADSSLIAG 787  
 QY 482 YSSVTVTSWPSIVATG-GTTT-----TPYSGSVTSSTKTTATASKTSTTRSG 531  
 DB 788 YGSLTGTAGVNSILTAYGSGTGTAKHESDLITGYGSTGTAGDSSLIAGVSTGTAG 843  
 RESULT 21  
 GUXI\_ASPAC STANDARD; PRT; 540 AA.  
 AC 059843;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, last sequence update)  
 DT 10-OCT-2003 (Rel. 42, last annotation update)  
 DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellulobiohydrolase I) (1,4-  
 DE beta-cellulobiohydrolase I) (Beta-glucanancellulobiohydrolase I).  
 GN CBHI.  
 OS Aspergillus aculeatus.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillius.  
 OX NCBI\_TaxId=5053;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=F-50;  
 RA Takada G., Kawaguchi T., Sumitani J., Arai M.;  
 RT "Cloning, nucleotide sequence, and transcriptional analysis of  
 RT Aspergillus aculeatus No.F-50 cellobiohydrolase I (cbhi) gene.";  
 RL J. Ferment. Bioeng. 85:1-9(1998).  
 CC -1- FUNCTION: The biological conversion of cellulose to glucose  
 CC generally requires three types of hydrolytic enzymes: (1)  
 CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)  
 CC Exocellulobiohydrolases that cut the disaccharide cellobiose from  
 CC the nonreducing end of the cellulose polymer chain; (3) Beta-1,4-  
 CC glucosidases which hydrolyze the cellobiose and other short cello-  
 CC oligosaccharides to glucose.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages  
 CC in cellulose and cellobiose, releasing cellobiose from the non-  
 CC reducing ends of the chains.  
 CC -1- SIMILARITY: Belongs to cellulase family C (family 7 of glycosyl  
 CC hydrolases).  
 CC -1- SIMILARITY: Contains 1 fungal-type cellulose-binding (CBD) domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB002821; BAA25183.1; -.  
 DR HSSP; P00725; IAZ6.  
 DR InterPro; IPR000254; CBD\_fungal.  
 DR InterPro; IPR008985; ConA\_like\_1ec\_g1.  
 DR InterPro; IPR001722; Glyco\_hydro\_7.  
 DR Pfam; PF00734; CBM\_1; 1.  
 DR Pfam; PF00840; Glyco\_hydro\_7; 1.  
 DR Prodom; PD001821; CBD\_fungal; 1.  
 DR Prodom; PD186135; Glyco\_hydro\_7; 1.  
 DR SMART; SM00236; fcbd; 1.  
 DR PROSITE; PS00562; CBD\_FUNGAL; FALSE\_NEG.  
 KM Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.



FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 540 EXOGLUCANASE 1.  
 FT DOMAIN 23 459 CATALYTIC.  
 FT DOMAIN 460 505 LINKER.  
 FT ACT SITE 506 540 CELLULOSE-BINDING (BY SIMILARITY).  
 FT ACT SITE 234 234 NUCLEOPHILE (BY SIMILARITY).  
 FT DISULFID 513 529 PROTON DONOR (BY SIMILARITY).  
 FT DISULFID 524 539 BY SIMILARITY.  
 FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 540 AA; 57099 MM; 002DFD8DF194D0 CRC64;

Query Match 5.6%; Score 156.5; DB 1; Length 540;  
 Best Local Similarity 23.3%; Pred. No. 0.0048;  
 Matches 85; Conservative 52; Mismatches 131; Indels 97; Gaps 19;

QY 223 GSAPATAVGSSGSCWQDAPELICYQ---SFWTGSFLANPDSRSRK-----D 269  
 DB 186 GAKYGT-----GYCDSQCPRLDKFIDQANIEGMEPSSTDVNAAGNHGSCPEMDIWE 239  
 QY 270 ANTLDSIHFPDEACDDSTFQPCSPRL-ANHKEVVDSPRSITLNDGLDSEAVAVG 328  
 DB 240 ANS-ISSAFYAHF---CDSVQQTCTGDTGCGTYSDDTDY-----SGTCDDPGCDFN 288  
 QY 329 RY-----PDITYNGNPMFLCTLAAEQLYDALYQWKOGL-EVTVSLDPEKAL 378  
 DB 289 PYFGNTPYFGPKTYDNSKPFYVTVQFTHDGT-----GTLFIRRLYVQNGVVI 341  
 QY 379 YSDAATGYSSSSSYSIYDAVKTADGFSIVETHAASNGMSSEQYK-SDEQI--- 434  
 DB 342 GNGPSYTYAASGNSITSEFCFAKTLF-GDTNVEFH---GGLSANGDALGQOMVLVS 396  
 QY 435 ---SARDLTWSAALLTANRRNSVVPASWGETSASVP---GTCNAF----- 476  
 DB 397 LMDHDHADMLM-----LSDYPTTSCASSPGVARGTCPTTGNATYVEAN 441  
 QY 477 --SAIGYSSV---TVTSMPSIVATGTTTAT--PIGSGSVTSTKTTASTKTTT 529  
 DB 442 YPMSYTYTYSIKRGTIANSTYSGTSSGSSSTTLTKASTTSSKTTTSTTSS 501  
 QY 530 SGMSL 534  
 DB 502 SSTNV 506

RESULT 22  
 ICER\_PSESX STANDARD; PRT; 1148 AA.

DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ice nucleation protein.  
 GN INAA.  
 OS Pseudomonas syringae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=317;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KCTC 1832;  
 RA Jung H.-C., Pan U.-G.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate  
 CC crystallization in supercooled water.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).  
 CC -1- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS  
 CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A  
 CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.  
 CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein  
 CC family.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1sb-rib.ch/announce/>  
 CC or send an email to [license@1sb-rib.ch](mailto:license@1sb-rib.ch)).

DR EMBL; AF013159; AAB66891.1; -  
 DR HSSP; P06620; 11NA.  
 DR InterPro; IPR000258; Ice\_nucleatn.  
 DR Pfam; PF00818; Ice\_nucleation; 57.  
 DR PRINTS; PR00327; ICENUCLEATN.  
 DR PROSITE; PS00314; ICE\_NUCLEATION; 38.  
 DR Ice nucleation; Repeat; Outer membrane.  
 FT DOMAIN 161 164 POLY-ALA.  
 FT SEQUENCE 1148 AA; 113811 MM; BA4019CF20FBE224 CRC64;

Query Match 5.6%; Score 155.5; DB 1; Length 1148;  
 Best Local Similarity 23.6%; Pred. No. 0.015;  
 Matches 137; Conservative 66; Mismatches 245; Indels 133; Gaps 27;

QY 32 SNEATYARTAILNNIGADG-----AWVGADSGIVVAPSTDNPD-----YYTWT 77  
 DB 200 SNETAGNHSDILAGYSTGTAGYSGTSGEDSLTAGYSGTQAQSGNLTAGYSGTGT 259  
 QY 78 --RDSGLVLTVDLFRNGDTSLSTIENYISQAIVQGSNBSGDLSS-----A 126  
 DB 260 AGSDSSLLIAGYSGTSGDSSSLTA--GYSGTQAQSGNLTAGYSGTGTAGVSSLLA 316  
 QY 127 GLGEPEFVNDERTAYTSGWGPQRDGPALBATAMIGFGWLLDNGYSTATDIWPLVRND 186  
 DB 317 GYSGTSGSDSLTAGYSGTQ-----TAGEGSN---LTAGYSGTGT-----AGSD 359  
 QY 187 LSYVAQYMQGTGYDLMEVNGSSFTTAV-QHRALVEGSAFATAVGSS-CSWCDQAPET 244  
 DB 360 SSLLA-----GYSGTSGSDSLTAGYSGTQAQSGNLTAGYSGTGTAGVSSLLIAG 413  
 QY 245 LCLYQGFMTGSFLANPDSRSRSGKANTL--LGSHTPDEACDDSTQPC-SPRALA 300  
 DB 414 YGSGTSGSDSLTAGYSGTQAQSGNLTAGYSGTGT-----AGADSSLLIAGYSGTSG 469  
 QY 301 NHKEVVDSPRSITLNDGL-----SDSEAVAVGRPEPTYNGNPMFLCTLA 347  
 DB 470 SSSSLTAGYSGTQAQSGNLTAGYSGTGTAGDSSLLA-GYSGTGT-----S 516  
 QY 348 AAEQLYDALY---QWKOGLSEVTVSLDPEKALYSDAATGYSSSSSYSIYDA--VK 402  
 DB 517 GSSSLTAGYSGTQAQSGNLTAGYSGTGTAGAAANLTGTYSGTGTAGHESFLIAGYS 576  
 QY 403 TPAQGFVSIV---ETHAASNGMSSEQYKSDGEQLSARDL-----TWSYALLTA 449  
 DB 577 TQTAGHKSILTAGYSGTGTARDSYLIAGYSGTGTAGSGSSLLIAGYSGTGTAYRSMULTA 636  
 QY 450 -----NNRRNSVVPASWGETSASVPGTCNAISAGTYSVTVTSMPSIYVAG-GTETT 502  
 DB 637 GYSGTGTARHSDLVTVGYSGTGTAG-----SNSSLTAGYSGTGTAGPKSLITAGYSGTGT 691  
 QY 503 ATP-----TSGSGSVTST-----KTTATASKTSTTT 528  
 DB 692 AGRSDVLVAGYGS-TSTAGYSSSLIAGYSGTGTAGYESTILT 731

RESULT 23  
 ICEA\_PANAN STANDARD; PRT; 1322 AA.  
 ID ICEA\_PANAN  
 AC P20459;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ice nucleation protein inaa.  
 GN INAA.  
 GN -----



OS Pantoea ananas (Erwinia uredovora).  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Pantoea.  
 CC NCBI\_TaxID=553;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90092494; PubMed=2599095;  
 RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;  
 RT "An ice nucleation active gene of Erwinia ananas. Sequence similarity  
 RT to those of Pseudomonas species and regions required for ice  
 RT nucleation activity."  
 RL FEMS Lett. 258:297-300(1989).  
 CC -|- FUNCTION: ice nucleation proteins enable bacteria to nucleate  
 CC crystallization in supercooled water.  
 CC -|- SUBCELLULAR LOCATION: Outer membrane (By similarity).  
 CC -|- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS  
 CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A  
 CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.  
 CC -|- SIMILARITY: Belongs to the bacterial ice nucleation protein  
 CC family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL, outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X17316; CA35194.1; -.  
 DR PIR: S07053; S07053.  
 DR HSSP: P06620; IINA.  
 DR InterPro: IPR000258; Ice\_nucleatn.  
 DR Pfam: PF00818; Ice\_nucleation; 69.  
 DR PRINTS: PR00327; ICENUCLEATN.  
 DR PROSITE: PS00314; ICE\_NUCLEATION; 49.  
 KW Ice nucleation; Repeat; Outer membrane.  
 FT DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.  
 FT SEQUENCE 1322 AA; 131094 MW; 89B0E24AA837039 CRC64;  
 SQ  
 Query Match 5.6%; Score 155; DB 1; Length 1322;  
 Best Local Similarity 22.0%; Pred. No. 0.02; Indels 154; Gaps 27;  
 Matches 125; Conservative 69; Mismatches 220;  
 QY 47 GADGAWVSGADSGIVASPTSDNPDYFTWTRDGLVLTVDLFRN-----GDTSL 99  
 DB 315 GADSLIAGVSTQTAGEESTQTAGVSTQTAKG-----SDTLTAVGSTGTAGDGS-- 366  
 QY 100 STEENYISAQAIYVQGINPFG-----DISG-----AGLGEPEKN 134  
 DB 367 SLIAGVSTQTAGDSSLTAGVSTQTAKGSDLTAGVSTGTSGADSLIAGVSTQTA 426  
 QY 135 VDEATVYSGWRPORDPALRATMIGFGWLNGTSTATDVTWVLVNDISYVAQW 194  
 DB 427 GEESTQTAGVSTG-----TAKGSD--LTAGVSTGT-----AGDSSLIAGV 469  
 QY 195 N-----QTGYLMEERVSSFFTLAVOHRALVEGSAFATAVSGSCWCDSPAELLCTYLS 250  
 DB 470 STQAQAGSUL-----TAGVSTGTAGYSSLIAG-----YGST-----QT 505  
 QY 251 FWTGSGFLANPDSRSGKADANTLL--GSHTFPDEACADSDTFQCPSPALANHKVVD 307  
 DB 506 AGVSTLTAGVSTQTQANESDILITGGYST--AGANSSLIAGVSTQTASYNVLT 561  
 QY 308 S-FRSYTLND-----GLSDEAVAVGVPEDTYNGPWILC-----TLAAE 350  
 DB 562 AGVSTGTARBSGLTLAGVSTGTAGSDSSIIAGVSTGTAGADSSIIAGVSTQTAGYN 621  
 QY 351 QLYDALYQMDKQSGLEVTVDLSLDFKALYSDAATGTATSSSSSTYSIVDAKTFADGFS 410  
 DB 622 SILTAGI-----GSTQTARBSGLTLAGVSTGTAGADSSIIAGVST-----TQTAGYN 670  
 QY 411 IV-----ETHAASNGS--MSROYDKSD-----GROLGA--RDLTWSYALLLTA 449

DB 671 ILTAGVSTQTAGEGSDLTLAGVSTGTAGDSSLIAGVSTQTASVHSSLIAGVSTQTA 730  
 QY 450 NNRNVSVPASMGTSASVPTCAATSAIGTVSVTVSWPSIVATG-GTTT-----503  
 DB 731 --QEGSVLTGVTGSTAG-----ADSSLIAGVSTQTAGYNILTAGVSTQTAGERSD 783  
 QY 504 TPTGSGSVTSTKTTATATASKTSTTTRSG 531  
 DB 784 LTTGVTGTAGADSSLIAGVSTQTAG 811  
 RESULT 24  
 ID ICEN\_PSESY STANDARD; PRT; 1200 AA.  
 AC P06620;  
 DT 01-JUN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ice nucleation protein.  
 GN INAZ.  
 OS Pseudomonas syringae (pv. syringae).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 CC NCBI\_TaxID=321;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5203;  
 RA Green R.L., Warren G.J.;  
 RT "Physical and functional repetition in a bacterial ice nucleation  
 RT gene."  
 RL Nature 317:645-648(1985).  
 RN [2]  
 RP SUBCELLULAR LOCATION.  
 RX MEDLINE=92297969; PubMed=2520825;  
 RA Lindow S.B., Lahue E., Govindarajan A.G., Panopoulos N.J., Gies D.;  
 RT "Localization of ice nucleation activity and the icec gene product in  
 RT Pseudomonas syringae and Escherichia coli."  
 RL Mol. Microbe Interact. 2:262-272(1989).  
 RN [3]  
 RP 3D-STRUCTURE MODELING OF 490-535.  
 RX MEDLINE=93360260; PubMed=8355267;  
 RA Kajava A.V., Lindow S.B.;  
 RT "A model of the three-dimensional structure of ice nucleation  
 RT proteins."  
 RL J. Mol. Biol. 232:709-717(1993).  
 CC -|- FUNCTION: ice nucleation proteins enable bacteria to nucleate  
 CC crystallization in supercooled water.  
 CC -|- SUBCELLULAR LOCATION: Outer membrane.  
 CC -|- DOMAIN: CONTAINS 122 IMPERFECT REPEATS OF THE CONSENSUS  
 CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A  
 CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.  
 CC -|- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE  
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.  
 CC -|- SIMILARITY: Belongs to the bacterial ice nucleation protein  
 CC family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL, outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X03035; CA26837.1; -.  
 DR PIR: A24405; SNPSO.  
 DR PDB: IINA; 31-OCT-93.  
 DR InterPro: IPR000258; Ice\_nucleatn.  
 DR Pfam: PF00818; Ice\_nucleation; 61.  
 DR PRINTS: PR00327; ICENUCLEATN.  
 DR PROSITE: PS00314; ICE\_NUCLEATION; 40.

KW Ice nucleation; Repeat; Outer membrane; 3D-structure.  
 FT DOMAIN 176 1151 OCTAPEPTIDE PERIODICITY.  
 FT STRAND 492 495  
 FT TURN 496 497  
 FT STRAND 498 501  
 FT STRAND 501 507  
 FT TURN 507 511  
 FT STRAND 512 513  
 FT STRAND 514 518  
 FT STRAND 524 527  
 FT TURN 528 529  
 FT STRAND 530 533  
 SQ SEQUENCE 1200 AA; 118587 MW; E3BEAD4BEA67ABE0 CRC64;  
 Query Match 5.6%; Score 154.5; DB 1; Length 1200;  
 Best Local Similarity 22.9%; Pred. No. 0.019;  
 Matches 144; Conservative 69; Mismatches 254; Indels 163; Gaps 29;  
 QY 9 LSGLVCTGLANVSKRATLDSWLSNEATVAFATILNNIGADGAWGAGDSGIVVAPSTD 68  
 DB 207 IAGYSGTGTAG-----SDSWL-----VAGYSTQTAGDSDLTGTAGYSTGTARBSNL 254  
 QY 69 NNDYFTTW--RDGLVLTVDLFRNCDTSLSTENTISAQAIYQGISNPGDLSG- 125  
 DB 255 TAGYSGTGTAGSDSLIAGYSTGTSGGSSLTAA--GYGSTQTAGBSNLGTAGYSTGT 311  
 QY 126 -----AGLGEPEKFNDETAYTSGMGRPO--RDGPALRA-----TAMIGFG 163  
 DB 312 AGSDSLIAGYSTGTSGGSSLTGTAGYSTGTQAGBSNLGTAGYSTGTAGVDSLIAGTG 371  
 QY 164 QWL-----LDNGYSTATDVIWPLVRNDLSVYAQYMNQGYD 200  
 DB 372 STQSGSDALTAGYSTGTQAGBSNLGTAGYSTGT-----AGSDSLIA-----GYG 419  
 QY 201 LMEVNGSFFITAV-QHRLVBSGAFATAVGSS-CSKCDSPAPILLCTLOSTWTSFPL 258  
 DB 420 STQTSGSDSLTAGYSTGTQAGBSLTGTAGYSTGTAGVDSLIAGYSTGTSGSSALT 479  
 QY 259 ANFDSRSRGKDNATL--LGSITFPDEACDSTPOPC-SPRALNHKEVDSFRSIYT 314  
 DB 480 AGYSGTGTAGBSNLGTAGYSTGT-----AGADSLIAGYSTGTSGSSSLTAGYSTGT 535  
 QY 315 LNDGL-----SDSEAVNAGRYPEDTYNGNPFLLCTLAARQLYDALY---Q 358  
 DB 536 AREGSFLTAGYSTGTAGADSLIA-GYSTGT-----SGSSSLTAGYSTGT 582  
 QY 359 WPKQGLEVTVDVSLDFKALYSDAATGYSSSSSTYSYDA--VKTADFVSIYETHA 416  
 DB 583 TAQQGSVLTSYGYSTGTAGAAASNLTTGYSTGTAGHSFIIAGYSTGTQAGHKSIL--T 639  
 QY 417 ASNGSMSEQYDKSD-----GQQLSA--RDLTWSYALLTA-----NNRN 454  
 DB 640 AGYSGTGTARBSDLIAGYSTGTAGSSSLIAGYSTGTASTRSLWLTAGYSTGTARH 699  
 QY 455 SVVPASWGETSASVPGTCAATSAIGTYSSVVTWSPSIVATG-GTTTATP-----TGS 508  
 DB 700 SLLVTCYGSTGTAG-----SNSSLIAGYSTGTAGAKRSILTAGYSTGTQAGERTSLVAGY 754  
 QY 509 GSVTSTST-----KTTATASKTSTTT 528  
 DB 755 GS-TSTAGYSSSLIAGYSTGTAGYSTTLLT 783  
 RESULT 25  
 ICEN\_PANAN STANDARD; PRT; 1034 AA.  
 AC ICEN\_PANAN 047879;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ice nucleation protein inau.  
 GN INAU.  
 OS Pantoea ananae (Erwinia uredovora).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Pantoea.  
 OX NCBI\_TaxID=553;  
 RN (1) -  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KUN-3;  
 RX MEDLINE=94264407; PubMed=7764866;  
 RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;  
 RT "Cloning and sequencing of an ice nucleation active gene of Erwinia  
 uredovora";  
 RL Bioeci. Biotechnol. Biochem. 58:762-764(1994).  
 CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate  
 CC crystallization in supercooled water.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -1- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE  
 CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE  
 CC PERIODICITY IS SUPERIMPOSED.  
 CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE  
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.  
 CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein  
 CC family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; D14992; BAA03636.1; -  
 DR PIR; JC2143; JC2143.  
 DR HSSP; P06620; 11NA.  
 DR InterPro; IPR000258; Ice\_nucleatn.  
 DR Pfam; PF00818; Ice\_nucleatn.  
 DR PRINTS; PR00327; ICENUCLEATN.  
 DR PROSITE; PS00314; ICE\_NUCLEATION; 34.  
 KW Ice nucleation; Repeat; Outer membrane.  
 FT DOMAIN 162 993 OCTAPEPTIDE PERIODICITY.  
 SQ SEQUENCE 1034 AA; 103378 MW; FA22523D333EAD CRC64;  
 Query Match 5.5%; Score 151.5; DB 1; Length 1034;  
 Best Local Similarity 21.9%; Pred. No. 0.024;  
 Matches 127; Conservative 60; Mismatches 210; Indels 183; Gaps 25;  
 QY 47 GADGAWGAGDSGIVVAPSTDNPDYFTWTDGLVLTVDLFRN-----GTSLL 99  
 DB 299 GADSLIAGYSTGTQAGBSLTGTAGYSTGTQAGK-----SDLTAGYSTGTAGDSDS 350  
 QY 100 STIENTISAQAIYQGISNPGS-----DLSSG-----AGLGEPEKN 134  
 DB 351 SLIAGYSTGTQAGBSDLTAGYSTGTQAGKSDLTGTAGYSTGTAGADSLIAGYSTGT 410  
 QY 135 VDETAYTSGWGRPPORDGPALRTATAMIGQWLLDNGYSTATDVIWPLVRNDLSVYAQY 194  
 DB 411 GESTGTAGYSTGT-----TAQKSD-----LTAGYSTGT-----AGDSSLIA--- 450  
 QY 195 NOTGYDLMEEVNGSFFITAVQHRLVBSGAFATAVGSSCSKCDSPAPILLCT--LQSTW 252  
 DB 451 ---GYGSTGTAGBSDLTAGYSTGTQAGKSDLTGTAGYSTGTAGYSSSLIAGYSTGT 507  
 QY 253 TGSFILLNPDSSRSRGKDNATL--GSITFPDEACDSTPOPCSPRALNHKEVDS- 308  
 DB 508 YGSLTLTAGYSTGTQAGNESDLITGYSTGT-----AGANSLIAGYSTGTASTRSLWLTAG 563  
 QY 309 FRSIYTLNDGLSDSEAVNAGRYPEDTYNGNPFLLCTLAARQLYDALYQWPKQGLEVT 368  
 DB 564 YGSLTGTARBG-SDLTAA-----GYGSTGT 586  
 QY 369 DVSIDPEFKALYSDAATGYSSSSSTYS--YDAVKTFADGVSIVETHAASNGSMSEY 426  
 DB 587 QEN-----SDLTGTAGYSTGTAGDSSLIAGYSTGTQAGYHSLI--TAGYSTGTQAG 635  
 QY 427 DKSD-----GEQLSAR--DLTWSYALL 446

```

Db      636 ERSDDLTTGCGSTAGADSSLIAGYSGTQAGYNSILTAGYSGTQTAQNSDLTTGCGST 695
Qy      447 LTNANRNSVVPASWGETSASVPGTCAA-----TSAIGTSSVTATSW-PSIV 494
Db      696 STINGESSLI--AGYSGTQTAASKSTLMAGYSSQTAREOSSLTAGYSGTSMAGYDSSLI 753
Qy      495 ATGGTTTAA-----TPTGSGSV-TSTSKTTATASKTSTTT 528
Db      754 AGYSGTQTAGYSGTTLTAGYSGTQTAHSSLTLAGYSGTAT 793

RESULT 26
MSB2_YEAST
ID      MSB2_YEAST      STANDARD;      PRT; 1306 AA.
AC      P23334;
DT      01-OCT-1993 (Rel. 27, Created)
DT      01-OCT-1993 (Rel. 27, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      MSB2 protein (Multicopy suppressor of bud emergence 2).
GN      MSB2 OR YGR014W.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX      NCBI_TaxID=4932;
RN      (1)
RP      SEQUENCE FROM N.A.
RX      MEDLINE=92383951; PubMed=1514328;
RA      Bender A., Pringle J.R.;
RT      "A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence
RL      defect."
RT      Yeast 8:315-323 (1992).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=S288C;
RX      MEDLINE=97435481; PubMed=9290212;
RA      Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT      "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT      chromosome VII."
RL      Yeast 13:1077-1090 (1997).
CC      -1- P1M: O-glycosylated in the Ser/Thr-rich regions (Probable).
CC      -1- SIMILARITY: SOME, TO YEAST HKR1.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC      or send an email to license@ebi.ac.uk).
CC      -----
DR      EMBL; M77354; AAA34798.1; -
DR      EMBL; Z72799; CAA96997.1; -
DR      PIR; S25370; S25370.
DR      Germodline; 141326; -.
DR      SGD; S0003246; MSB2.
DR      GO; GO:0005887; C:integral to plasma membrane; IMP.
DR      GO; GO:0005034; F:osmosensor activity; IMP.
DR      GO; GO:0000283; P:establishment of cell polarity (sensu Sacch. .; IGI.
DR      GO; GO:0006970; P:response to osmotic stress; IMP.
KW      Transmembrane; Glycoprotein; Repeat.
FT      DOMAIN 698 816 7 17 AA TANDEM REPEATS.
FT      REPEAT 698 714 1.
FT      REPEAT 715 731 2.
FT      REPEAT 732 748 3.
FT      REPEAT 749 765 4.
FT      REPEAT 766 782 5.
FT      REPEAT 783 799 6.
FT      REPEAT 800 816 7.
SO      SEQUENCE 1306 AA; 133114 MW; 67D5D984D5CA4A6D CRC64;

Query Match 5.5%; Score 151.5; DB 1; Length 1306;
Best Local Similarity 22.9%; Pred. No. 0.034;

```

```

Matches 134; Conservative 83; Mismatches 240; Indels 129; Gaps 24;
Qy      25 ATLDMSINEATVARTAI-LNNIGADG-----WVSGADSGIIVASPTDNPDTYTRD 79
Db      267 APLQTSSESSFTTASALPVSSTVDGSSASPVVMSAAGQALSSSTDNPTMSEFTSL 326
Qy      80 SGLVLTLYDLFRNGTSLSTIENTISAQAIYQGISNPSGDLSSAGLGEPEFVNDER 139
Db      327 S-----TEVD-----GSPVVS--STVSALLSAFLQTSNPSFIYSPVSFVPSGSSD--- 373
Qy      140 YTGSGWRPQDPALATAMIGFGWLNDNGYSTADTDIYWPVLRNLDLSVAQYMNQTC 199
Db      374 -----VASSSTANVSSSFSDIPPTSTSGSV-----SVAQASALAF 412
Qy      200 DLWEVNGSS-----FFTIAVQHRALVEGSAFPAVAGSSGSCWD-----SOAPEIL 245
Db      413 QSTTEVYGASASGTWMSLSLSTSLQSTTL-DSSSLSSASSSSDLIDYGVSRASIPLS 471
Qy      246 CYIQSIWTGSPFI-----ANPDSRSGKD-ANTLGSIHFPDPAACDDSTFOGCFRAL 300
Db      472 ASQASTSSSFVSFVSFVSFVSFVSQSSDVASTSAPSVS-----SSFYTSIQAG 522
Qy      301 NHEVVDPRSIYTLNDGLSDSEAVAVGRY-----PEDTYNGN-----PWFICTIAAEOL 352
Db      523 SSMTNPSSSTIYSSSTGSSSESAATAATLSSGSSITMAGLQSPSTSLSESQA 582
Qy      353 -----YDALYOMDKQSLLEVTVSLDFPKALYSDA-----TGTVSSGS 391
Db      583 TSTSAVLASSSVGTSTPTTAGAGTAASLSLSTSAETASQVSGSTTALQTSFPASS 642
Qy      392 -----STYSIYDAVTFADGFVSIYETIAANGSMSEQYD-KSDEQLSARD 438
Db      643 TTEGSETSSQGFSTSVLVQMPESLSEFSPSQGTTQNMASASSSQYTIISGTILSQVSD 702
Qy      439 LTNVYALLTANRRRSV--VPASWGETSASVPGTCAATSAIGTSS-----YTV 487
Db      703 TSVSYT--TSSSVSQVSDTPVSY-TTSSSVSQVSDTPVSYTTSSSVSQVSDTPVSY 758
Qy      488 TSWPSIVATGCTTTTATPTGSGSVSTSKTTATASKTSTYTRSGMS 533
Db      759 TTSSSVSQVSDTPVSYTTSSSVSQVSDT-----SVPTSSRSVS 800

RESULT 27
MANB CALSA
ID      MANB CALSA      STANDARD;      PRT; 1331 AA.
AC      P22533;
DT      01-ANG-1991 (Rel. 19, Created)
DT      01-JUL-1993 (Rel. 26, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Beta-mannanase/endoglucanase A precursor (includes: Mannan endo-1,4-
DE      beta-mannosidase A (EC 3.2.1.78) (Beta-mannanase) (Endo-1,4-
DE      mannanase); Endo-1,4-beta-glucanase (EC 3.2.1.4) (Cellulase)).
GN      MANA.
OS      Caldociellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC      Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC      Caldicellulosiruptor.
OX      NCBI_TaxID=44001;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=93119139; PubMed=1476429;
RA      Gibbs M.D., Saul D.J., Luthi E., Bergquist P.L.;
RT      "The beta-mannanase from 'Caldociellum saccharolyticum' is part of a
RT      multidenton enzyme."
RL      Appl. Environ. Microbiol. 58:3864-3867 (1992).
RN      [2]
RP      SEQUENCE OF 1-346 FROM N.A.
RX      MEDLINE=91247819; PubMed=2039230;
RA      Luthi E., Jasmat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
RT      "Cloning, sequence analysis, and expression in Escherichia coli of a
RT      gene coding for a beta-mannanase from the extremely thermophilic
RT      bacterium 'Caldociellum saccharolyticum'."
RL      Appl. Environ. Microbiol. 57:694-700 (1991).

```

CC -1- FUNCTION: DEGRADATION OF HEMICELLULOSES, THE SECOND MOST ABUNDANT  
 CC POLYSACCHARIDES IN NATURE. CONTAINS TWO CATALYTIC DOMAINS WITH  
 CC MANNANASE AND ENDOGLUCANASE ACTIVITIES.  
 CC -1- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic  
 CC linkages in mannans, galactomannans, glucomannans, and  
 CC galactoglucomannans.  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC MISCELLANEOUS: This enzyme is most active at pH 6 and 80 degrees  
 CC Celsius.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO CELLULASE FAMILY  
 CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, BELONGS TO CELLULASE FAMILY  
 CC J (FAMILY 44 OF GLYCOSYL HYDROLASES).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L01257; AAA71887.1; -  
 CC EMBL: M36063; AAA72861.1; -  
 CC PIR: A48954; A48954.  
 CC HSSP: Q06851; INBC.  
 CC InterPro: IPR001956; CBD\_3.  
 CC InterPro: IPR008965; Cellul\_1.  
 CC InterPro: IPR001547; Glyco\_hydro\_5.  
 CC Pfam: PF00942; CBM\_3; 2.  
 CC Pfam: PF00150; Cellulase; 1.  
 CC ProDom: PD001947; CBD\_3; 2.  
 CC PROSITE: PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 CC PROSITE: PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 CC Hydrolyase; Glycosidase; Cellulose degradation; signal;  
 CC KM  
 CC Multifunctional enzyme.  
 CC FT CHAIN 1 41 POTENTIAL.  
 CC FT DOMAIN 42 1331 BETA-MANNANASE/ENDOGLUCANASE A.  
 CC FT DOMAIN 42 325 CATALYTIC (MANNANASE ACTIVITY).  
 CC FT DOMAIN 326 361 PRO/SER/THR-RICH (PT BOX).  
 CC FT DOMAIN 362 518 SUBSTRATE-BINDING (POTENTIAL).  
 CC FT DOMAIN 519 564 PRO/SER/THR-RICH (PT BOX).  
 CC FT DOMAIN 565 720 SUBSTRATE-BINDING (POTENTIAL).  
 CC FT DOMAIN 721 780 PRO/SER/THR-RICH (PT BOX).  
 CC FT DOMAIN 781 1331 CATALYTIC (ENDOGLUCANASE ACTIVITY).  
 CC FT ACT\_SITE 162 162 PROTON DONOR (BY SIMILARITY).  
 CC FT ACT\_SITE 257 257 NUCLEOPHILE (BY SIMILARITY).  
 CC FT CONFLICT 338 338 T -> P (IN REF. 2).  
 CC FT CONFLICT 340 346 TPPTPT -> ROHQHQ (IN REF. 2).  
 CC SEQUENCE 1331 AA; 146892 MW; FPCAS1BB8D8F0D CRC64;  
 CC -----  
 CC Query Match 5.4%; Score 151.5; DB 1; Length 1331;  
 CC Best Local Similarity 21.1%; Pred. No. 0.034;  
 CC Matches 102; Conservative 60; Mismatches 161; Indels 161; Gaps 20;  
 CC -----  
 CC 84 LKTLVDFRGRDPSLSTENIYISAQIVGGINSPS-GDSSGAGLGEKFNVDERTAYTG 142  
 CC 3 LKTKI--RKRWLSVCTVFLNLIETIANVTILPKVGAATSDGV---VKIDTSTLIG 55  
 CC -----  
 CC 143 S-----WGRPDGPALRATAMIGFGW-----LLDNGYTSATDITVPLVRNDLSYV 190  
 CC 56 TNAHCHMYRRLD-----TALRGIRMGKNSVRVUSNIGR-----SEVANITISLSRS 91  
 CC -----  
 CC 191 AQYWGOTGYLWEEVNGSSFTTAVOHRALVEGSAFATVAVSSCWCSDQAPILCYLOS 250  
 CC 92 -----WTKIPA-----SEVANITISLSRS 109  
 CC -----  
 CC 251 FWGSGFLIANPDSRSRGKANDATLIGSIHTFDEPAACDSTFGQSPALANHEKVDSEFR 310  
 CC 110 LGFRKAILLEVHDTTGVEDG-----AAC-----SLA--QAVEYWK 142  
 CC -----  
 CC 311 SIYTLNDGLSDSEAVAVGARYPEDTYVNGN--PWFLCTIAAEQYDALYO-----W 359

DB 143 EIKSVLDGNEDEFTIINIGNEP---YGNNNYQWVNDTKAIALRADRGKHTIMVADPNW 199  
 QY 360 DKQSGLEVTVDSLDPEKA-----LYSDAATGYSSSSSTYSIVADVTPADGFSIV- 412  
 DB 200 GQWSTMTKDNOSIMEADPLRNLYVSIHMYGV---NTASVYEYIISFVDKGLPIVY 255  
 QY 413 --ETHAASNGSMSEQYDKSDGEQLSARDLTW-----SYAALLTANRRNSVVPASWGET 464  
 DB 256 GERGHGHTGDPDDEALIVRAYQKIGLFSWNGKNSVYGVLDVNMNDPNNPTPMGQ- 314  
 QY 465 SASVVGTCALATSALITGSVYTSWPSIVAVGCTTTATPFGSGSVTSKTTATAASKT 524  
 DB 315 -----WKTWALGTSSTPTPTS-----TVPPTPTPTPTPTVTAIPTPTPVST 360  
 QY 525 STTT 528  
 DB 361 PATS 364  
 -----  
 CC RESULT 28  
 CC FIG2\_YEAST STANDARD; PRT; 1609 AA.  
 CC ID FIG2\_YEAST  
 CC AC P25653;  
 CC DT 01-MAY-1992 (Rel. 22, Created)  
 CC DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 CC DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 CC DE Factor induced gene 2  
 CC GN FIG2 OR YCR089W OR YCR89W OR YCR1102.  
 CC OS Saccharomyces cerevisiae (Baker's yeast).  
 CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 CC NCBI\_Taxid=4932;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=92397594; PubMed=1523889;  
 CC RA Wilson C., Grisanti P., Frontali L.,  
 CC RT "The complete sequence of a 6146 bp fragment of Saccharomyces  
 CC cerevisiae chromosome III contains two new open reading frames."  
 CC RL Yeast 8:569-578 (1992).  
 CC CC  
 CC -1- FUNCTION: Required for efficient mating.  
 CC -1- INDUCTION: By mating pheromones.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X59720; CAA42254.1; -  
 CC PIR: S25345; S25345.  
 CC Germonline: 136986; -  
 CC SGD: S0000685; FIG2.  
 CC DR GO: GO:0009277; Cell wall (sensus Fungi); IDA.  
 CC DR GO: GO:0000753; P:cellular morphogenesis during conjugation w. .; IMP.  
 CC DR GO: GO:0000755; P:cytogamy; IMP.  
 CC SEQUENCE 1609 AA; 166049 MW; 7D66AD7F85A7B852 CRC64;  
 CC -----  
 CC Query Match 5.4%; Score 150; DB 1; Length 1609;  
 CC Best Local Similarity 20.2%; Pred. No. 0.056;  
 CC Matches 124; Conservative 92; Mismatches 211; Indels 186; Gaps 32;  
 CC -----  
 CC 61 VVASPTDNDPVYTYTTRDGLVLTVDLFR---NGDTSLSLTENIYISAQIVGIGIS 116  
 CC 312 VLTSNSID-PSLFTTSEVSSTQLSLNRSKSEYTFASISTFTFGTDSATSLIDPIS 370  
 CC -----  
 CC 117 NPSGDLSSGAGLGEKFN-----VDERTAYGS-----WGRQGRGPAALRATAMIGFGQ 164  
 CC 371 SVGSTASSFVGISTANFSTQNSNYPBESTASSSOYQDWSSESL--PLSQTT----- 421  
 CC -----  
 CC 165 WLIDN-----GYTSTAD-----IYW-PL-----VANDLSYVA 191

```
Db 422 WVVINTNTGSGVSTTSPAVSVTATKTVGVTETVTCWCPLOTKSOAIGVSSSISVP 481
QY 192 QYNNQGTGYDLMEEVNGSSFFTIYAVQHRALVEGSAFATAVGSS--CSWCDSDAPEILCYLQ 249
Db 482 QASSFSG-----SSILSNSSTSLAASNNV-----PESTASSSSQYQDMSSSLP-----LSQ 528
QY 250 SFWTGSITLANPDSRSGKANTLGSIRH-----FDP 282
Db 529 TTV-----VVINTNTGSGVSTTSPAVSVTATKTVGVTETVTCWCPLOTKSOAIGISS 584
QY 283 EAACDDSTFQPCSPRALA-NHKEVVD-SFRSITVTLNDGLSDSEAVAVGRYPEDTYNG-- 338
Db 585 STSATGTSPSSSLTIGISTQLSDATFKGTETINHHM-TESTST---TEPTREGTS 640
QY 339 NPWFLECT---LAAEQLYDALYQWMDKQSLVTVDSLDFEKLALYSDAATGYSSS--- 390
Db 641 DSFYLTCTSEVNLASLSSEYFN-FSSSEGSTATINSTVTF-----GSITSKYPTSTSVN 692
QY 391 -----SSTSSIVAVATFADGFVSIVETHAASNG--SMSQYDKSDGEO----- 433
Db 693 PTEASQHVSSSVNSLTPFTSNSTETIYAVINIKHTSNKDYSLTTQTLKTSKQTLVLST 752
QY 434 -----LSARDLTW-----SYAL-----LTANNRRNSVVP 458
Db 753 VTTTNGAATEYTTWCASSIAVTTISIKYTLVLTTEVCHSECTPVITVSATSTIP 812
QY 459 ASWGETSASSVPGTCAATSAIG-TYSSVTVTSMPISVATGTTTATPTSGSVTSSTKT 517
Db 813 LL-STSSSTVLSSTVEGAKNPASAEVTINT-QVATSEATSTSTQVATSATATASE 868
QY 518 TATRSKSTSTTRS 530
Db 869 SSTTSQVSTASSET 881

RESULT 29
YG46_YEAST STANDARD; PRT; 507 AA.
ID YG46_YEAST STANDARD; PRT; 507 AA.
AC P53301;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 52.8 kDa protein in BUB1-HIP1 intergenic region.
GN YG189C OR G7553.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
ON NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA MEDLINE=97279231; PubMed=9133739;
RX Arroyo J, Garcia-Gonzalez M, Garcia-Saez M.I, Sanchez-Perez M,
  Nombela C, Garcia-Gonzalez M, Garcia-Saez M.I, Sanchez-Perez M,
  "DNA sequence analysis of a 23,002 bp DNA fragment of the right arm
  of Saccharomyces cerevisiae chromosome VII.",
  Yeast 13:357-363(1997).
RL -1 SIMILARITY: SOME, TO YEAST UTR2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; Z72974; CA97215.1; -
CC EMBL; X99074; CA67525.1; -
CC PIR; S64507; S64507.
CC HSSP; P23904; 1AJK.
CC Germonline; 141501; -
CC SGD; S0003421; CRH1.
```

```
DR GO; GO:0009277; C:cell wall (sensu Fungi); IDA.
DR GO; GO:0000131; C:incipient bud site; IDA.
DR InterPro; IPR008985; Cons like 1ec_g1.
DR InterPro; IPR00757; Glyco hydro_16.
DR Pfam; PF00722; Glyco_hydro_16; 1.
KW Hypothetical protein
FT DOMAIN 63
FT DOMAIN 301
FT DOMAIN 345
FT DOMAIN 387
FT DOMAIN 467
SQ SEQUENCE 507 AA; 52757 MW; 7D7B61F57AE942C CRC64;

Query Match 5.3%; Score 147; DB 1; Length 507;
Best local similarity 20.8%; Pred. No. 0.019;
Matches 125; Conservative 75; Mismatches 143; Indels 258; Gaps 29;

QY 17 LANVISKRATLDWLSNENAVTARTALINNGADGAVNSGADSGIVVAPSTDPDYFTW 76
Db 6 LITVLSASSLSLTFAAESTAT-----ADSTTAASSTASCNP----- 42
QY 77 TRDSGLVLTVDLFRNG--DTSLSTLENIYSAQAIYQGISNPGDLSGAGLEPKF 133
Db 43 -----LKT-----TGCTPDAL-----ATSFSDPSSSKWFTDLGHAEIKY 80
QY 134 NDEIATYSGWRPQDGPALRATAMIGFGQWILDNGYSTATDVIWPLVRNDLSVAYQ 193
Db 81 GSDGLSMTLA--KRDNPGLKSNFYIMYK-----LEVILKA 115
QY 194 WNOTGYDLMEEVNGSSFFTIYAVQHRALVEGSAFATAVGSSCSWCDSDAPEILCYLOS--- 250
Db 116 ANGTGI-----VSSP-----YLOSDDL 132
QY 251 -----FWTG--SPIIANPDSRS-----GKDANTLGSIHFPPEAACDSTF- 291
Db 133 DEIDIEWGGDNPQPSNFFSKDITTYDRGEFHGVDTP--DKFNITLDMAMDKTWY 190
QY 292 -QPCSPALANHKEVVDSEFRSITYTLNDGLSDSEAVAVGRYPEDTY-NGNP--- 340
Db 191 LDSESVRLSN-----TSSEG-----YPCQPMYLMGMWAGDPPDNA 227
QY 341 -----WFLCTLAABOLY-DALYQWMDKQSLVTVDSLDFEKLALYSDAATG--TSSSS 391
Db 228 AGTIEM--ACGETNYNDAPF-----TMYIEKIVTVDYSGKKYTGDDOS 269
QY 392 STYSSIVDAVKTPADGFVSIVETHAASNGSMSEQYDK-----SDGEOLASRDY---T 440
Db 270 GSWESIE--ADG-----GSIVGRYQAGEDFAVLANGSISSTSSST 311
QY 441 WSYAALLTANNRRNSVVPASWGETSASVPGTCAATSAIGTVSSV---TTSMPISIVA 495
Db 312 VSSASASTVSSSVSTVSSASSTVSSSVSSSVSSSVSSSVSSSTPSSSTATSSKTLAS 371
QY 496 TGGTT-----TTATPTGSGVTSSTKTAATASKTS-----TTRSGM 532
Db 372 SSVTTSSISSIFEKQSSSKKTVASSSTSEISSTSKTPATVSTTRSTVAPPTQSSV 431
QY 533 S 533
Db 432 S 432

RESULT 30
ALAI_CANAL STANDARD; PRT; 1419 AA.
ID ALAI_CANAL STANDARD; PRT; 1419 AA.
AC O13386;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Agglutinin-like protein ALAI precursor (Agglutinin-like adhesin).
GN ALAI OR ALSS.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
```

CC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98053977; PubMed=9393828;  
 RA Gaur N.K., Klotz S.A.;  
 RT "Expression, cloning, and characterization of a Candida albicans  
 gene, ALA1, that confers adherence properties upon Saccharomyces  
 cerevisiae for extracellular matrix proteins.";  
 RL Infect. Immun. 65:5289-5294(1997).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.  
 CC -1- PTM: N-glycosylated and O-glycosylated (Potential).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
 CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
 CC  
 CC EMBL: AF025429; AAB88883.1; -  
 DR PIR: T30531; T30531.  
 DR InterPro: IPR008440; Candida\_ALS.  
 DR Pfam: PF05792; Candida\_ALS; 1.  
 KW Cell adhesion; Glycoprotein; Repeat; Signal.  
 FT SIGNAL 1 17  
 FT CHAIN 18 1419 AGGLUTININ-LIKE PROTEIN ALA1.  
 FT DOMAIN 399 404 POLY-THR.  
 FT DOMAIN 408 418 POLY-THR.  
 FT DOMAIN 437 441 POLY-THR.  
 FT DOMAIN 673 676 POLY-SER.  
 FT DOMAIN 687 690 POLY-SER.  
 FT DOMAIN 700 703 POLY-SER.  
 FT DOMAIN 719 724 POLY-SER.  
 FT DOMAIN 749 752 POLY-SER.  
 FT DOMAIN 787 791 POLY-SER.  
 FT DOMAIN 869 872 POLY-SER.  
 FT DOMAIN 875 883 POLY-SER.  
 FT DOMAIN 901 911 POLY-SER.  
 FT DOMAIN 1216 1221 POLY-SER.  
 FT CARBOHYD 665 665 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1301 1301 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1326 1326 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 1419 AA; 149635 MW; 249F3F688A9D5B6 CRC64;  
 Query Match 5.1%; Score 142; DB 1; Length 1419;  
 Best Local Similarity 19.9%; Pred. No. 0.16;  
 Matches 134; Conservative 89; Mismatches 210; Indels 242; Gaps 34;  
 OY 26 TLDSWLSNEATVARTALNIGADGAVSGADSGIVVASTPNPDYFY--WTRDSGLV 83  
 DB 440 TQDFW--SESFTSTTTITNSL-----KQTD--VIAREP--NHPVTTFEWS--ESFAT 486  
 OY 84 LKTLVD-----LRNGDTSLSTENTYISAQAVQSGN--PSGDLSSGAGIGEE--- 111  
 DB 487 TETITKPEGTDSVIVREPNPTVTTFEWSSEIATTEITINGPEG--TDSVIAREHPN 544  
 OY 132 ----XFNVDETAVTGS--WGRPORD-----GPALRATAMIGFQMLDNGYSTAT 176  
 DB 545 TVTTTFWMSYSVATTEITINKPEGTDSVIVKEPYNPTVTTFE-----W--SESYATTEI 597  
 OY 177 DIWMP-----LYR--NDLSYVAQVYNGQGYDLMEEWGSSFFTTAVQHRALVEGSAFA 227  
 DB 598 ITNGPEGTDSVIVREPNPTVTTFEWSSEIATTEITINGPEG--TDSVIAREHPN 544  
 OY 228 TAVGSSCSWGDQAPETLCYQSWTGSFLLANDSRSGKANDTLGSIHTPPEACD 287  
 DB 657 TALESSDSNISSSAHO-----SSSYEQSSSIVIGLSSSSDIPSSD 697  
 OY 288 DSTFQPCSPALANHKEVVDVFRSIVYTLNDQLSDSEAVAVAGRPEDTYVYNGNPFWFLCTLA 347

DB 698 ----MFSSS-----STGLTSSSESTVSSYSD--SSSSIESSTUS 731  
 OY 348 AAEQ-----LYDALYQMDKQGS--LEVTVY-----SLD----- 373  
 DB 732 SSDRCSSSISDITTSFWDSSSDIESTISITWSSSIDAQSHLYQSVNSISTISQELSSSS 791  
 OY 374 -----FFRALYSDAATGYSSSSSTRY-----SSIVDAVK 402  
 DB 792 EESSTPATDALVSDPSSSLSDTSYPSSTRISSSDDPPTIAGSDLSISFISTVYE 851  
 OY 403 TPADGVSVIVETHAAS--NSMSGEQYKSGDEQLSADLTWSYALLTANNRRNSVYP--- 458  
 DB 852 ISSDS--VSILSPDPSFSSSSSLNSDSSSPSSDQSD-----ILTSSSFLLVPSFS 903  
 OY 459 -----ASWGETSASVPGTCA-----ATSAIGTY 482  
 DB 904 LSSSSSLSTLPHYVNSTYTHASESSSSVASSMSASGANDTYLSEDTTSSIGTD 963  
 OY 483 SSVT-----VTSWPS-----IVATGCTTTATPTGS--GSVTSKRTA 519  
 DB 964 SSTVTFGRDNGDCIVTGMPSSSIDSEQTSDVTTTSSFFVASSTPTSAEQSITDNPIDS 1023  
 OY 520 TASKSTTTRSGMGL 534  
 DB 1024 SQTSASSSTKSSVS 1038  
 RESULT 31  
 ID XYND\_CELFI STANDARD; PRT; 644 AA;  
 AC P54865;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Endo-1,4-beta-xylanase D precursor (EC 3.2.1.8) (Xylanase D) (XYLD).  
 OS Cellulomonas faml.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Micrococcineae; Cellulomonadaceae; Cellulomonas.  
 OX NCBI\_TaxID=1708;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=221;  
 RX MEDLINE=9424155; PubMed=8170399;  
 RA Millward-Sadler S.J., Poole D.M., Hentisat B., Hazlewood G.P.,  
 RA Clarke J.H., Gilbert H.U.;  
 RT "Evidence for a general role for high-affinity non-catalytic  
 RT cellulose binding domains in microbial plant cell wall hydrolases.";  
 RL Mol. Microbiol. 11:375-382(1994).  
 CC -1- FUNCTION: Endo-acting xylanase which displays no detectable  
 CC activity against polysaccharides other than xylan. Hydrolyzes  
 CC glucosidic bonds with retention of anomeric configuration.  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic  
 CC linkages in xylans.  
 CC -1- PATHWAY: xylan degradation.  
 CC -1- SIMILARITY: Belongs to cellulase family G (family 11 of glycosyl  
 CC hydrolases).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
 CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
 CC EMBL: X76729; CAA54145.1; -  
 DR PIR: I40712; I40712.  
 DR PDB: 1E5B; 25-MAY-01.  
 DR PDB: 1E5C; 25-MAY-01.  
 DR PDB: 1HEH; 10-MAY-01.  
 DR PDB: 1HEJ; 10-MAY-01.  
 DR PDB: 1XBD; 21-JUL-99.



DR PDB: 2XBD, 21-JUL-99.  
 DR InterPro: IPR001919; Bac\_cellose-bind.  
 DR InterPro: IPR008965; Cellul bind.  
 DR InterPro: IPR008985; Cona\_like\_1ec.g1.  
 DR InterPro: IPR001137; Glyco\_hydro\_11.  
 DR InterPro: IPR002509; Polyac\_deacet.  
 DR Pfam: PR00553; CBM\_2, 2.  
 DR Pfam: PR00457; Glyco\_hydro\_11; 1.  
 DR Pfam: PF01522; Polyac\_deacet; 1.  
 DR PRINTS; PR00911; GLHYDRLASE11.  
 DR SMART; SM00637; CBD\_11; 2.  
 DR PROSITE; PS00776; GLYCOSYL\_HYDROL\_F11\_1; 1.  
 DR PROSITE; PS00777; GLYCOSYL\_HYDROL\_F11\_2; 1.  
 KW Xylan degradation; Hydrolase; Glycosidase; Signal; Repeat;  
 KM 3D-structure.  
 FT SIGNAL 1 43 POTENTIAL.  
 FT CHAIN 44 644 ENDO-1,4-BETA-XYLANASE D.  
 FT DOMAIN 44 230 CATALYTIC.  
 FT DOMAIN 231 245 LINKER ("HINGE") (GLY-RICH BOX).  
 FT DOMAIN 246 644 2 X 88 AA APPROXIMATE CBD-LIKE REPEATS.  
 FT REPEAT 246 333 1.  
 FT REPEAT 557 644 2.  
 FT DOMAIN 337 350 LINKER ("HINGE") (PRO-THR BOX).  
 FT DOMAIN 337 350 LINKER ("HINGE") (GLY-RICH BOX).  
 FT DOMAIN 231 238 POLY-GLY.  
 FT DOMAIN 241 245 POLY-GLY.  
 FT DOMAIN 548 558 POLY-GLY.  
 FT ACT\_SITE 126 126 NUCLEOPHILE (BY SIMILARITY).  
 FT ACT\_SITE 216 216 PROTON DONOR (BY SIMILARITY).  
 SQ SEQUENCE 644 AA; 66581 MW; 56B045CC6E0B1820 CRC64;

Query Match 5.1%; Score 141.5; DB 1; Length 644;  
 Best Local Similarity 20.7%; Pred. No. 0.059;  
 Matches 142; Conservative 76; Mismatches 238; Indels 229; Gaps 35;

QY 28 DSWLSNATYARXAI--LNNIGADGAWVSGADGIVVASPS---TDN-----PDYRYT- 75  
 DB 3 DSEATRTTRRRRLQALTLGILAAALVAGH--LAAASPAALAVTSNTTGTHDGYTSF 59  
 QY 76 WTRDSGLVLTVDLFRNGDTSLSTIENYISAQAVOGISNPSGDLSSGAGEPRFNV 135  
 DB 60 WTDSPGSV---SMDLNGGGGTWRSNTGNFVAG---KMGSTGRKTVSVSG---QFNP 108  
 QY 136 DERYTSSKWRPQDGPALATAMIGQWMLDNGYSTATD-----IYPLV 183  
 DB 109 SRNAVLLTYLWMTQ--SPLVEYIYVDSWGYRPTGTFMGVTYSDGYDIYRTQVKNPFI 166  
 QY 184 RNDLSYVAQYWN-----QTG-----YDLWE-----EYNGSSF 211  
 DB 167 EGDSSSTYQYVWVQQRKTGTTSNGHFDAMASKNGLGHNYMAMATRGYSSGSSSI 226  
 QY 212 TIAVQHRALVEGSAFA---TAVGSSCWCDSPAPEILCYLOSP-----WTGSF 256  
 DB 227 TVS-----EGSGGGGGSDPTGGGGSTGCSVTAIRAEMSDRNVTVSVSGSSAWTVNL 279  
 QY 257 ILANFDSRSRGKAN-----TLGSIHTP-----DPEAACDSTFPPCS 295  
 DB 280 ALNGSQITQASWNAVTVGSGSTRITVPNGSGNTFGVTVMKGSSTPAATCAGSGGTAT 339  
 QY 296 PRALANHKEVVDPRSRY--TLNDG-----LSDSEVAVAGRYPDRTVYVNGP 340  
 DB 340 PTPTPTPTTPPOSAGVGLTFDDGPNGTGTTNOIISTLTQYGVAT--VPTTGNAAQGNP 398  
 QY 341 WFLCTLAABQLYD-----ALYQMD-----KQGSLEVTVSIDPF 375  
 DB 399 SLW-----QAYKNAGVQIGNHSHDHPHVMNSQSDMQSLRTTQQAICQTAGVTPLF 451  
 QY 376 KALY--SDAATGYSSS-----SSTYSIVDAVKTPADGFVSIV- 413  
 DB 452 RPPYGEENATLRQVESLGLREIITWVDSDPMNNAASQIRQOASRLTNGQIIMHMPA 511  
 QY 414 -THAASGMSSEQYDKSDGEOISARDLTWSYAALLTANNRNSVVPAS--WETSASSVPG 471

DB 512 ATVOALPGLI-----QDLRSRLCTGHISSTG-----RAVAPSSAGGGGGGGCTG 558  
 QY 472 TCAATSAIG-----TY-----SSVTY-----SPSIVATGTTTATPT 506  
 DB 559 SCVSVAVRGEMADRFRVTVSVSGSWVTLGLANGGQVOSWMNALNTSSGTVARPN 618  
 QY 507 GSGS---VT-STSKTATASKSTT 527  
 DB 619 GSGNSFGVTFYKNGSSATPGATCAT 643  
 RESULT 32  
 HKRL\_YEAST  
 ID HKRL\_YEAST STANDARD; PRT; 1802 AA.  
 AC P41809;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, last sequence update)  
 DT 15-MAR-2004 (Rel. 43, last annotation update)  
 DE Hansenula WRAKII killer toxin-resistant protein 1 precursor.  
 GN HKRL OR YDR420W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=YNN 295;  
 RX MEDLINE=94156857; PubMed=8113191;  
 RA Kasahara S., Yamada H., Mio T., Shiraori Y., Miyamoto C.,  
 RA Yabe T., Nakajima T., Ichishima E., Furuchi Y.;  
 RT "Cloning of the Saccharomyces cerevisiae gene whose overexpression  
 RT overcomes the effects of Hm-1 killer toxin, which inhibits  
 RT beta-glucan synthesis.";  
 RT J. Bacteriol. 176:1488-1499 (1994).  
 CC -I- FUNCTION: Could regulate beta-glucan synthesis. Overexpression  
 CC -I- provides resistance to Hm-1 killer toxin.  
 CC -I- SUBCELLULAR LOCATION: Type I membrane protein (probable).  
 CC -I- PTM: Could be O-glycosylated in the serine/threonine-rich domain.  
 CC -I- SIMILARITY: SOME, TO YEAST MSB2.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, S69101; AAB30051.1; -.  
 DR GenBank; 140912; -.  
 DR SGD; S0002828; HKRL.  
 KW Glycoprotein; Transmembrane; Repeat; Signal.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 1802 HANSENULA WRAKII KILLER TOXIN-RESISTANT  
 FT TRANSMEM 1486 1506 PROTEIN 1.  
 FT DOMAIN 23 1478 SER/THR-RICH.  
 FT DOMAIN 453 788 12 X 28 AA TNDEN REPEATS OF S-[AV]-[P]-  
 FT V-A-V-S-S-T-Y-T-S-S-S-P-A-A-I-S-S-T-  
 FT Y-T-S-S-P.  
 FT REPEAT 453 480 1 (APPROXIMATE).  
 FT REPEAT 481 508 2.  
 FT REPEAT 509 536 3.  
 FT REPEAT 537 564 4.  
 FT REPEAT 565 592 5.  
 FT REPEAT 593 620 6.  
 FT REPEAT 621 648 7.  
 FT REPEAT 649 676 8.  
 FT REPEAT 677 704 9.  
 FT REPEAT 705 732 10.  
 FT REPEAT 733 760 11.  
 FT REPEAT 761 788 12.  
 FT CARBOHYD 24 N-LINKED (GLCNAC. .) (POTENTIAL).



FT CARBOHYD 1252 1252 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1293 1293 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1342 1342 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1400 1400 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1802 AA; 188890 MW; E344CA6469785A24 CRC64;

Query Match 5.1%; Score 141; DB 1; Length 1802;

Best Local Similarity 21.4%; Pred. No. 0.26;  
 Matches 127; Conservative 78; Mismatches 243; Indels 146; Gaps 25;

QY 15 TGLANVSRKATLDSKLSNEATVATFALLNMGADGAWGSGDGI--VVASPTD-NPD 71  
 DB 234 TFEFSNTVS-----FENVEEYAMSKQLSSYSSSTVSGGSEADTKTSSPITSFSSS 289  
 QY 72 YFTYTW-----TRDSGLVLUKLYVDLFRNGDITSLSTIE---NYISQAQIV 112  
 DB 290 YSGTITSTESSESRVAVGVRPSITQTSTIDSFMSSELYLSTYDLSAGNYPDQELIYD 349  
 QY 113 ----QGISNPSGDLSSGAGLGEPEKFNVDETAYVGSWGRPQRDGPALRATAMIGQWLLD 168  
 DB 350 RPASTSTATSTSSAAGGVRESNTFAVSSIST----- 382  
 QY 169 NGYTSTATIV-----WPLVNDLSYVAQYNNQGYDIWE---EVNGSSPFTIA 214  
 DB 383 NFIVSASDPTVVSSTSTNTVPYSSVHSTFVHATSSSTYISSILYSSPSSLASVSHFGVA 442  
 QY 215 VQHRALVESAFAFATVAGSSCWCDS-----QAPETLCYLOGFMTGS-----FILA 259  
 DB 443 PEPASATISFSSVPAVAVSTYSSPSASVVPVAVASPPVAVAVSTYSSPSAPALISS 502  
 QY 260 NFDSSRSGKDANTLLGSIHTPDPEA-ACDDSTP--OPCSPRALA---NHKEVVDSEFRSI 312  
 DB 503 YTTSSSS---APAVASSTYSSPSAPALISSSTYSSPSAPAVASSTYSSPSAPALISS 559  
 QY 313 YTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABQLYDALYQMDKQGLVETDVL 372  
 DB 560 YT---SSSPAPAV---SSTYSSP-----SAPALISSSTYSSPSAPAVAVSTY 603  
 QY 373 DFFKALYSDAATGTVSS-----SSTYSSIVDAKTFADGVSIVETHAANGSMSE 425  
 DB 604 SSPSA-PAIISTYSSPSAPAVAVSTYSSPSAPALISSSTYSSPSAPVPA---VSS 657  
 QY 426 YKSIDSEQLSARDLTVSYALLTANNRRNSVVPASGNETSASVPTCAATSAIGTVS-- 483  
 DB 658 YTSSP-----SAPALISSSTYSSPSAPVAVASSTYSS--PSAPALISSSTYSS 705  
 QY 484 ----SVTVSWPSIVATGCTTTATPTGSGVTS--TSKTTATASKTSITTT 530  
 DB 706 APVAVSTYSSPSAPALISSSTYSSPSAPAVAVSTYSSPSAPALISSSTYSS 759

RESULT 33

ICEN\_XANCT STANDARD; PRT; 1567 AA.  
 AC P18127;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ice nucleation protein.  
 GN INAX.  
 OS Xanthomonas campestris (pv. translucens).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 NCBI\_TaxId=343;  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=X56S;  
 RX MEDLINE=91080859; PubMed=2259339;  
 RA Zhao J., Orser C.S.;  
 RT "Conserved repetition in the ice nucleation gene inax from  
 RT Xanthomonas campestris pv. translucens.";  
 RL Mol. Gen. Genet. 223:163-166(1990).  
 CC -I- FUNCTION: Ice nucleation proteins enable bacteria to nucleate

CC crystallization in supercooled water.  
 CC -I- SUBCELLULAR LOCATION: Outer membrane (By similarity).  
 CC -I- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS  
 CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A  
 CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.  
 CC -I- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE  
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.  
 CC -I- SIMILARITY: Belongs to the bacterial ice nucleation protein  
 CC family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC DR EMBL; X52970; CAA37140.1; -.  
 DR HSSP; P06620; IINA.  
 DR InterPro; IPR000258; Ice\_nucleatn.  
 DR Pfam; PF00818; Ice\_nucleation; 81.  
 DR PRINTS; PR00327; ICNUCLEATN.  
 DR PROSITE; PS00314; ICE\_NUCLEATION; 57.  
 KW Ice nucleation; Repeat; Outer membrane.  
 SQ SEQUENCE 1567 AA; 152548 MW; CBB451D959ECAD63 CRC64;

Query Match 5.1%; Score 140.5; DB 1; Length 1567;

Best Local Similarity 23.1%; Pred. No. 0.23;  
 Matches 126; Conservative 68; Mismatches 229; Indels 123; Gaps 25;

QY 47 GADGAMV-----SGASGIVVASPSND-----NPDYTYWT--RDGSLVLUKTLV 88  
 DB 725 GADSTLIAGYSTQTSIGSDSSSLTAGYGTQARKGSDVTAGYSTGTAGADSTLIAGYGS 784  
 QY 89 DIFRNDGTSLSTIENYISAQIVQGISNPSGDLSSG-----AGLGEPEKFNVD 139  
 DB 785 TQTSIGSDSSSLTA---GYSTQARKGSDITAGYGSTGTAGADSTLIAGYGSTQTSIGSD 841  
 QY 140 YTGSGWRPQ--RDGPALRATAMIGFQW-----LLDNGYTSTATDIYVPLVNDLSY 190  
 DB 842 LTAGYSTQTAAREGSDVTA---GYSTGTAGADSTLIAGYGSTQTA---AGSDSSSL 891  
 QY 191 AQYMN-QTYDLMEEVNGSSFTIANQHRALVBSGAFATVAGSSCWCMSQAPETLCYIQ 249  
 DB 892 AGYGTQTAARKGSDVTAGYSTGTAGADSTLIAGYGSTQTSIGS---DSSLTAGYGSTQ 946  
 QY 250 SEWTSFPIIANFDSRSGKDANTLLGSIHTPDPEAACDDSTPQCSPRALANHKEVVDSE 309  
 DB 947 TARKGSDMTAGYSTGTAGADSTLIAGYGSTQTSIGSDSSSLTAGYGSTQTAAREGSDV 1006  
 QY 310 RSIYTLNDGLSDSEAVAVGRYPEDTYNGNPMFLCTLAABQLYDALY--QMDKQGLSE 366  
 DB 1007 GS-----TGTAGADSTLIAGYGS-----TGTAGSDSSLTAGYGSTQTAARKG 1048  
 QY 367 VTDVSLDFFKALYSDAATGTVSSS-----STYSSIVDAKTFADGVSIVETHAANG 419  
 DB 1049 --DVTAGY-----GSTGTAGADSTLIAGYGSTQTSIGSDSSSLTAGYG-----STYTA 1094  
 QY 420 GS-MSEQYDKSD-----GEOLSA---RDLTWSYALLTANNRRNSVVPASWGE 463  
 DB 1095 GSDITAGYSGTGTAGADSSSLIAGYGTQARKGSDNTTAGYSGTQA--REDSSLTAGYGS 1152  
 QY 464 TSASSVPTCAATSAIGTVSSVTVSWPSIVATG-GTTTATPTGSGVTSISKTTATAS 522  
 DB 1153 TSTAG-----HDSLIAGYGTQTAGYNSILTTGYGSGTQTAQ--SSSLTAGYGSTTAG 1205  
 QY 523 KTSTTT 528  
 DB 1206 YDSTLT 1211

RESULT 34

```

SP96.DICDI
ID _SP96.DICDI STANDARD; PRT; 600 AA.
AC PI4328;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Spore coat protein Sp96.
GN Dictyostelium discoideum (Slime mold).
OS Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=90067962; PubMed=2587278;
RA Fosenau K., Loomis W.F.;
RT "Sequence of the Dictyostelium discoideum spore coat gene SP96."
RL Nucleic Acids Res. 17:9489-9489(1989).
CC -1- SUBCELLULAR LOCATION: Outer layer of the coat matrix and interspersed matrix.
CC -1- PTM: PHOSPHORYLATED AND FCOSYLATED. MAY BE PHOSPHOGLYCOSYLATED, MAY CONTAIN GLCNAC-ALPHA-1-P-SER RESIDUES.
CC -1- SIMILARITY: Contains 4 preprope motif repeats.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).
-----
CC CC
CC EMBL; X16491; CAA34508.1; -.
DR PIR; S07638; S07638.
DR GlycosultEDB; PI4328; -.
DR DictyBase; DDB0185032; cotA.
DR InterPro; IPR007643; Dict_spore_N.
DR InterPro; IPR003645; FOIN.
DR Pfam; PF04562; Dict_spore_N_1.
DR SMART; SM00274; FOLN; 6.
KW Glycoprotein; Phosphorylation; Repeat; Sporulation.
FT REPEAT 185 197 PRESORE MOTIF 1.
FT REPEAT 221 233 PRESORE MOTIF 2.
FT REPEAT 298 310 PRESORE MOTIF 3.
FT REPEAT 395 407 PRESORE MOTIF 4.
SQ SEQUENCE 600 AA; 59589 MW; 61EA6D02B5F1071 CRC64;

Query Match 5.0%; Score 138.5; DB 1; Length 600;
Best Local Similarity 29.3%; Pred. No. 0.085;
Matches 44; Conservative 32; Mismatches 73; Indels 1; Gaps 14.

QY 380 SDAATGTSSSSSTYSIVDAKVTADGFVSIETHASNGSMSEQYKSDOEQLSARDL 439
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 447 SSSASGSSAAASSPSSSAASSPSSSAASSPSSSAASSPSSSAASSPSSSAASSSA 506
   ::::::::::::::::::::
QY 440 TWSYAALLTA-NNRNSVVPASWGETSASVPGTCATSAIGTSVTVTSNPISVATGG 498
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 507 PSSASSSSSAPSSSAASSSSSSASSSASNAATTAATTATAATTATAATTATAATTA 566
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 499 TTTTATPTGSGSVSTSKTKTATAKTISTTT 528
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 567 TTTAATTATTATAATTATAATTATAATTATAATTATTATTATTATTATTATTATT 596
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 35
VGIX_HSVB VGIX_HSVB STANDARD; PRT; 797 AA.
AC P28968;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Glycoprotein X precursor.
GN

```

```

OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1);
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1.";
RL Virology 189:304-316 (1992).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-eb.ch/announce/
CC or send an email to license@isb-eb.ch).
-----
CC
CC EMBL, M86664; AAB02506.1; -.
DR PIR, H36802; VGBEX1.
DR InterPro: IPR007110; Ig-like.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 22
FT CHAIN 23 797 GLYCOPROTEIN X.
FT DOMAIN 23 465 SER/THR-RICH.
FT TRANSMEM 766 790 POTENTIAL.
FT CARBOHYD 590 590 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 797 AA; 80342 MW; 50C9ED9211FE5B82 CRC64;
-----
Query Match 5.0%; Score 138; DB 1; Length 797;
Best Local Similarity 32.0%; Pred. No. 0.13;
Matches 47; Conservative 21; Mismatches 59; Indels 20; Gaps 4;
-----
OY 400 AVKTEPAQPFYSIVETHTAASNGSMSEQYDKSDGEQLSARDLTWSYALLTANNRNSVPPA 459
Db 14 AVSLVIALGSTITTETTTSSSSSTSSGSGSTGTTNSSSPSTTSPT--TSSSPSTHTS 71
OY 460 SWGHTS-----ASSVPGTCATSAIGTYSV-TVTSWPSIVATGTTTATPTGSG 509
Db 72 SPSSTSTGSSSTATSSASPAVSTASITSIPTSTETTTTTPASTTPTTTTAAPTTAA 131
OY 510 SV-----TSTSKTTATASKTSTTT 528
Db 132 TTTAVTTAATSAETTTATATATATSTPT 158
-----
RESULT 36
WCS3_YEAST STANDARD; PRT; 556 AA.
ID ID WCS3_YEAST Q1215;
AC AC Q1215;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cell wall integrity and stress response component 3 precursor.
GN WCS3 OR YOL105C OR HRS56.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96076631; PubMed=7502582;
RA Vandenbol M., Durand P., Portetelle D., Hilger F.;
RT "Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV
RT including the Y1-H3 retrotransposon, the SUI1(+) frameshift
RT suppressor gene for tRNA-Gly, the yeast transfer RNA-Thr-1a and a
RT delta element.";
RL Yeast 11:1069-1075 (1995).
CC -1- SIMILARITY: Contains 1 WSC domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; 248149; CAA88155.1; --  
CC EMBL; 274847; CAA9123.1; --  
CC PIR; S51892; S51892.  
CC Germonline; 143527; --  
CC SCD; S0005465; MSC3.  
CC GO; GO:0004888; F:transmembrane receptor activity; IGI.  
CC GO; GO:0009408; P:response to heat; IGI.  
CC GO; GO:0007266; P:rho protein signal transduction; IGI.  
CC InterPro; IPR002889; MSC.  
CC Pfam; PF01822; MSC; 1.  
CC SMART; SM00321; MSC; 1.  
CC Cell wall; Transmembrane; Glycoprotein; Signal.  
CC SIGNAL 1  
CC CHAIN 39 556  
CC DOMAIN 39 132  
CC TRANSSEM 137 348  
CC TRANSSEM 385 405  
CC CARBOHYD 84 84  
CC CARBOHYD 367 367  
CC CARBOHYD 370 370  
CC CARBOHYD 473 473  
CC CARBOHYD 480 480  
CC SEQUENCE 556 AA; 58229 MW; DD37E27180001DA CRC64;  
Query Match 4.9%; Score 137; DB 1; Length 556;  
Best Local Similarity 27.7%; Pred. No. 0.096;  
Matches 48; Conservative 41; Mismatches 68; Indels 16; Gaps 5;

QY 371 SLDFPKAL--YSDAATGYSS-----SSSTYSIYDAVKTPADGFSIVETHAANGS 421  
DB 93 SVSFLSLTSTSDNCGTKSCGWPYMGCGSSYTN-----YVNAEFVSSVESSSMEGS 148  
QY 422 MSQOYKSDCEQLSADLTMSYALLTANNRNSVVPASGERSASVPCTAATATCT 461  
DB 149 -STSYMPSTTSSLSAQI--SSTTRRTSTMKSEMIATVSTTSSSTSSSTSS 205  
QY 482 YSEVTVMPSPSIATGTTTATPTGSGSVTSKTTATATSKTSTTRSGMSL 534  
DB 206 STTSSSTTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSS 258  
RESULT 37  
CH12\_RHIOL STANDARD; PRT; 542 AA.  
AC P29027;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE Chitinase 2 precursor (EC 3.2.1.14).  
GN CH12  
OS Rhizopus oligosporus.  
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;  
OC Rhizopus.  
OX NCBI\_TaxID=4847;  
RN (1)  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-38.  
RC STRAIN=Saio IFO 8631;  
RX MEDLINE=93054356; PubMed=1429462;  
RA Yanai K., Takaya N., Kojima N., Horiuchi H., Ohta A., Takagi M.;  
RT "Purification of two chitinases from Rhizopus oligosporus and  
RT isolation and sequencing of the encoding genes";  
RL J. Bacteriol. 174:7398-7406(1992).  
CC -1- FUNCTION: Probably involved in the apical growth and branching  
CC of fungal hyphae.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-  
CC acetyl-D-glucosamine polymers of chitin.

CC -1- SUBUNIT: Monomer.  
CC -1- SUBCELLULAR LOCATION: Secreted (Probable).  
CC -1- PTM: O-glycosylated.  
CC -1- SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl  
CC hydrolases).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D10158; BAA01022.1; --  
CC PIR; B47022; B47022.  
CC HSSP; P23472; 2HW.  
CC InterPro; IPR001223; Glyco\_hydro\_18.  
CC InterPro; IPR001579; Glyco\_hydro\_18A.  
CC InterPro; IPR005089; Glyco\_hydro\_18c.  
CC Pfam; PF03427; CBM\_19; 1.  
CC DR PF00704; Glyco\_hydro\_18; 1.  
CC DR PROSITE; PS01095; CHITINASE\_18; 1.  
CC Hydrolase; Glycosidase; Chitin degradation; Signal; Zymogen;  
CC Glycoprotein; Chitin-binding.  
CC SIGNAL 1 22  
CC CHAIN 23 446  
CC PROPER 447 542  
CC DOMAIN 23 312  
CC DOMAIN 313 354  
CC DOMAIN 355 406  
CC DOMAIN 407 446  
CC ACT SITE 166 166  
CC SEQUENCE 542 AA; 56528 MW; 3B8F17DA551FDDA7 CRC64;  
Query Match 4.9%; Score 136; DB 1; Length 542;  
Best Local Similarity 23.1%; Pred. No. 0.11;  
Matches 83; Conservative 50; Mismatches 146; Indels 80; Gaps 17;

QY 224 SAFATVSSGSCWCSQAPETLCYL-QSFWTGSFILANFDS-SRSGKANTLLGSIHTFD 261  
DB 12 SAEFLASTGVQAW-SHGPNVWYWGQNSGASNTQASIGTSCQVDAVILFLHVN 70  
QY 282 ----PE-----AACDDSTPQPCS-----PRALAHKEVVDSPRSIYTLNDGLDSEAVAVGR 329  
DB 71 VGGIPEINLSSACAGYFPNTQLSCPAVAGDIKKQDGVKVT-----LSLGAAGVYG 125  
QY 330 YPEDFYNGNPWFLCTLAABQLYDALYQWDKQSLVTVDSLDFPKALYSDAATGYSS 389  
DB 126 FTSQ-----AGQGFAPQITWMLFGGNSD-----TRPFDAYIDGVDLIEG 168  
QY 390 SSETYSIYDAVK-----TPADGFSIVETHA--SNGSMSEQYKSDG 431  
DB 169 SSTGYVAFNALRQKFSNFLIGAPQCFPPDALLGSVINSASFDVYVQFYNNYSATG 228  
QY 432 EQLSARDLTMSYALLTANNRNSV--VPASGERSASVPCTCAAT-----SAIGRY 482  
DB 229 SSEFND-TWDWAKTSPNKVKIMFTVPGSSTAAGSYVPSTLQITVPSIASRYSS 286  
QY 483 SSVTV--TSWP-----SIVATGTTTATPTGSGSVTSKTTATATSKTSTTT 528  
DB 287 GGVSVWDASQAKNNGFNSQLYSLVHSGG-STPPPPSSSATKTTTKTTATSKTITTT 344  
RESULT 38  
Y109\_YEAST STANDARD; PRT; 995 AA.  
ID Y109\_YEAST  
AC P40442;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hypothetical 99.7 kDa protein in SDL1 5' region precursor.  
GN Y1169C OR Y19402.07C.



QY 274 LGS1-----HTPPEAACDSTFOPCSPRALANH--- 302  
 Db 245 SNAVTHPCDTPSCQMTGRCGCTYSIDRXCCTCPD-GCDNPVR-----MCTVTFYCG 299  
 QY 303 KEVVDVFRS1LYL-----NDGLSDSEAVAVGR-YEDTYTNGNPNPFLCTLAALAEQLYDA 355  
 Db 300 GETIDT-KSPFTVVTGFLTNIDGTSTGLSEIKRFYVQGGKVINP----- 343  
 QY 356 LVQMDQGLS-----EVDVSLDFPKALYSDAATGYSSSSSTYSIYDAVTFADGV 409  
 Db 344 -----QSTIVGVSGNSITDSWCMACKSAFGD-----TNEFSKHQGMAGMAGLADGV 391  
 QY 410 -----SIVETHAASNGSMSEQYKSDGEQLSARDLTWGYAALLTANNRNSVVPASWGETSA 466  
 Db 392 LVMSLWDDHAS-----DVLMDSTYPT--NATSTTPGAKRGTCOD 429  
 QY 467 SSVPTGCAAT--SAITGYSSVTVTSMPSIVATGTT---TTATPTGSGSVTSTKTTAT 520  
 Db 430 SRPFTVESTYPAAVYISNIK-TGPLNSTFTGCTTSSSTTTTSTKSTSTSSSKTTT 488  
 QY 521 ASKTSTTTRSGMS 533  
 Db 489 V--TTTTTSSGSS 499

RESULT 40  
 PHP\_DROME STANDARD; PRT; 1589 AA.

AC P39769;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Polyhomeotic (proximal) chromatin protein.  
 GN PH-P.  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Imaginal discs;  
 RX MEDLINE=92146957; PubMed=1346609;  
 RA Decanilis M., Cheng N.S., Pierre D., Brock H.W.;  
 RT "The polyhomeotic gene of Drosophila encodes a chromatin protein that  
 RL shares polyene chromosome-binding sites with Polycomb.";  
 RL Gene Dev. 6:223-232(1992).  
 RN (2)  
 RP SEQUENCE OF 199-1584 FROM N.A.  
 RX MEDLINE=92039031; PubMed=1937015;  
 RA Deatrick J., Dally M., Randsholt N.B., Brock H.W.;  
 RT "The complex genetic locus polyhomeotic in Drosophila melanogaster  
 RL potentially encodes two homologous zinc-finger proteins.";  
 RL Gene 105:185-195(1991).  
 CC -1- FUNCTION: Binds to polyene chromosome. Seems to interact with  
 CC Pc. May interact with proteins already bound to promoter  
 CC complexes and may be a negative regulator of homeotic and  
 CC segmentation genes. Plays a role in regulating the expression of  
 CC other pair-rule genes such as eve, ftz, and H.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: Salivary glands.  
 CC -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.  
 CC -1- SIMILARITY: TO MOUSE EARLY DEVELOPMENT REGULATOR PROTEIN RAE-28.  
 CC -1- CAUTION: It is uncertain whether Met-1 or Met-9 is the initiator.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X63672; CAA45211.1; -  
 DR EMBL; M64750; -; NOT\_ANNOTATED\_CDS.  
 DR PDB; 1KW4; 05-JUN-02.  
 DR FlyBase; Fggn0004861; ph-P.  
 DR GO; GO:0016458; P:gene allencing; IGI.  
 DR InterPro; IPR001660; SAM.  
 DR Pfam; PF00536; SAM; 1.  
 DR SMART; SM00454; SAM; 1.  
 DR PROSITE; PS50105; SAM\_DOMAIN; 1.  
 KW Zinc-finger; Developmental protein; DNA-binding; Nuclear protein;  
 KW 3D-structure.  
 FT ZN FING. 1365 1387 C4-TYPE.  
 FT DOMAIN 1513 1577 SAM.  
 FT DOMAIN 74 80 POLY-GLN.  
 FT DOMAIN 411 450 GLN-RICH.  
 FT DOMAIN 494 520 GLN-RICH.  
 FT DOMAIN 619 650 GLN-RICH.  
 FT DOMAIN 775 960 GLN-RICH.  
 FT DOMAIN 1233 1290 SER/THR-RICH.  
 FT CONFLICT 254 254 MISSING (IN REF. 2).  
 FT CONFLICT 1415 1415 D -> A (IN REF. 2).  
 SQ SEQUENCE 1589 AA; 167297 MW; A6DF0CF9106E1B91 CRC64;

Query Match 4.8%; Score 133.5; DB 1; Length 1589;  
 Best Local Similarity 24.6%; Pred. No. 0.67;  
 Matches 86; Conservative 43; Mismatches 151; Indels 69; Gaps 15;

QY 214 AVQHRALVGSAPATVAVSSGSCWDSQAPELLQYLOSFWTGSFTLANPSSRSKQANTL 273  
 Db 982 ALQALASAGAIQFQAKPCTCSSSPSTSVITNQS-----STPLVTSSTVASIQQAQTO 1037  
 QY 274 LGS1HFD---PEACDDSTFOPCSPRALANKEVVDVFRS1YTLNDG-LSDSEAVAVGR 329  
 Db 1038 SAQVHQHQQILSATIATAGTQOQPGPSLPTTNPLTANSMNATVGLSLSTAPVTVS- 1096  
 QY 330 YEDTYTNGNPNPFLCTLAALAEQLYDALYQMDKQSLVVD-----VSLDFPK 376  
 Db 1097 -YTAIVATSSPOLVILSTASS-----GGGGSIPATPTKETPSPKGTATVPIGSPK 1147  
 QY 377 ALYS--DAATGYSSSSSTYSIYDAVKTFFADGVSVETHAASNGSMSEQYKSDGEQL 434  
 Db 1148 TPVSGKDTCTTPKSSTPATVASVSEASSSTGE-----ALSNG-----DASD----- 1188  
 QY 435 SARDLTWGYAALLTANNRNSV--PASWGETSASVPG-----TCAATSAITGYSS 484  
 Db 1189 --RSSTLSKGAATTPYSKQNAAVQPPS--TTPNSVSGKEPKLATCGSLTSAISTSTT 1243  
 QY 485 VTWTSMPSIVATGTTTATPTGSGSVTSTKTTATASKTSTTRSGMS 533  
 Db 1244 TTTTNGIGVART--TASTAVSTASTTTTSSGTFRTTCTSTTTTSSIS 1290

RESULT 41  
 XANP\_XANS2 STANDARD; PRT; 827 AA.

AC 060106;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Xanthomonas precursor (EC 3.4.21.101) (Xanthomonas aspartic  
 DE proteinase) (Xanthomonaspepsin) (Carboxyl proteinase) (XCP).  
 OS Xanthomonas sp. (strain T-22).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 NCBI\_TaxID=136420;  
 RN (1)  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=97058302; PubMed=8902622;  
 RA Oda K., Ito M., Uchida K., Shibano Y., Fukuhara K.-I., Takahashi S.;  
 RT "Cloning and expression of an isovaleryl peptidase in *in-sensitiv*  
 RL carboxyl proteinase gene from *Xanthomonas* sp. T-22.";  
 RL J. Biochem. 120:564-572(1996).

RN [2]  
 RP MUTAGENESIS OF ASPARTIC ACID AND GLUTAMIC ACID RESIDUES, AND ACTIVE  
 RP SITES.  
 RX MEDLINE=99419069; PubMed=10488127;  
 RA Oyama H., Abe S.-I., Ushiyama S., Takahashi S., Oda K.;  
 RT "Identification of catalytic residues of pepstatin-insensitive  
 RT carboxyl proteinases from prokaryotes by site-directed mutagenesis.";  
 RL J. Biol. Chem. 274:27815-27822(1999).  
 CC - CATALYTIC ACTIVITY: Cleavage of casein.  
 CC - COFACTOR: Binds 1 calcium ion per subunit (By similarity).  
 CC - SUBCELLULAR LOCATION: Secreted.  
 CC - SIMILARITY: Belongs to peptidase family S33.  
 CC - SIMILARITY: Contains 1 PKD domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.emb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: D83740, BAI2093.1, -.  
 CC MEROPS: S53.002; -.  
 DR InterPro: IPR000601; PKD.  
 DR InterPro: IPR007280; PPC.  
 DR Pfam: PF00082; Peptidase\_S8, 1.  
 DR Pfam: PF00801; PKD, 1.  
 DR Pfam: PF04151; PPC, 1.  
 DR SMART: SM00089; PKD, 1.  
 DR PROSITE: PS50093; PKD, 1.  
 KM Hydroxylase; Protease; Serine protease; Zymogen; Signal;  
 KM Calcium-binding.  
 FT SIGNAL 1 23 POTENTIAL.  
 FT PROPEP 24 237 REMOVED IN MATURE FORM.  
 FT CHAIN 238 635 XANTHOMONALISIN.  
 FT PROPEP 636 827 REMOVED IN MATURE FORM.  
 FT DOMAIN 635 722 PKD.  
 FT ACT\_SITE 312 312 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 316 316 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 544 544 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT METAL 585 585 CALCIUM (VIA CARBONYL OXYGEN)  
 FT METAL 586 586 CALCIUM (BY SIMILARITY).  
 FT METAL 603 603 CALCIUM (VIA CARBONYL OXYGEN)  
 FT METAL 603 603 CALCIUM (BY SIMILARITY).  
 FT METAL 605 605 CALCIUM (BY SIMILARITY).  
 SQ SEQUENCE 827 AA; 83706 MW; 21A33C4C683DBC9F CRC64;  
 Query Match 4.8%; Score 132; DB 1; Length 827;  
 Best Local Similarity 19.9%; Pred. No. 0.35;  
 Matches 122; Conservative 72; Mismatches 221; Indels 197; Gaps 27;  
 QY 16 GLANVTSKRAVTLDSWLSNEATV-----AFVAILNNGIGAGWVSGADSGIVVASPT 67  
 DB 205 GLQNVSVKHLHMYHDEDTVPGRPNVGTQAAVAHHPDFAIYVGGSS-----LPAAT 260  
 QY 68 DNPDPYFTWTRDSGLVUKTVLVL--FANGDPSLSTENISAAIYQGISNPSGDSLSG 125  
 DB 261 NTAVGIIITW---GSIQTVDLNSFTSG--AGLATVNSITL-----KVGSG 301  
 QY 126 AGLEPKFNVDETAFTYSGWGRPORD--GPALRATAMIGF-----GQWLLDNG-----Y 171  
 DB 302 TFPANDPDSN-----GEMSLDSQDIVIGIAGGVQQLIFYISANGDSSSGITTDGITASY 354  
 QY 172 TSTATDVLWPLVRNDLSYVAGYMNQGYDLMEVNGSSFFTTIAVQHRALVEGSAFATAVG 231  
 DB 355 NRAVTDNIAKLINVSLGDEDTAAQSGTQAADD-----ALFQGAVAAGQGFSTIASG 405  
 QY 232 SSGCWCSQAPEILCYQSFWTSGFIILANPSSSGSDANTLLGSHITPDEAACDSTF 291  
 DB 406 -----DAGVYQWST---DPTSG----- 419

QY 292 QPCSPALANHKKEVDSFESITYLTNDGLSDSEAVAGRRYPEDTYNGNPWF-----L 343  
 DB 420 ---SPGVANASAGTVKIDLTHYVSESPASSPYIYQVGGTTLST--SGTTWSGETVWNEG 474  
 QY 344 CTIAAAB-----QLYDALYQMDKXGSELVTVSLDF--FKLLYSPAATGYTS 388  
 DB 475 SAIAFPQGDNNORLMATGCGVSLYEAPSPQSSVSTKRVGPDLAFDAASSGALIVYA 534  
 QY 389 SSSSTSSSIYDAVATPADGFVSIVETHAASNGSM---SEQYDKSDGEQLSARDLTWSYA 444  
 DB 535 GSTEQVGTSLASPLVGAFAPI---ESAANNAIGPASKFYQAFPTQSLSLHDTV----- 587  
 QY 445 ALLITANN-----RNSVVPASW-----GETSASSVPG-T 472  
 DB 588 ---SGNNQYQSHGTYATGFEATGFGSPDICKLNTVYAQNMWTVGGGGGTNAPVANSF 644  
 QY 473 CATSAIGTYSSTVYWSVPIVA-----TCGTTTATPT-----GSGSV--TSNKT 518  
 DB 645 VATTGLVAATFTDSDSDGSIASHAWTFDGGSTSTATSPBHTYSAAGTYSVAETVIDNAG 704  
 QY 519 ATASKTSTTTRS 530  
 DB 705 ATSTKTSSTVTS 716  
 RESULT 42  
 FLO1 YEAST STANDARD; PRT; 1537 AA.  
 ID FLO1 YEAST  
 AC P32766;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Flocculation protein FLO1 precursor (Flocculin 1).  
 OS FLO1 OR YAR050W.  
 GN Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94262325; PubMed=8203162;  
 RA Watari U., Takata Y., Ogawa M., Sahara H., Koshino S., Omela M.-L.,  
 RA Aitakinen U., Jaatinen R., Penttilae M., Keranen S.;  
 RT "Molecular cloning and analysis of the yeast flocculation gene FLO1.";  
 RL Yeast 10:211-225(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=95249563; PubMed=7731988;  
 RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,  
 RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,  
 RA Storms R.K.;  
 RT "The nucleotide sequence of chromosome I from Saccharomyces  
 RT cerevisiae.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).  
 RN [3]  
 RP PRELIMINARY SEQUENCE FROM N.A.  
 RX MEDLINE=93289821; PubMed=8511970;  
 RA Teunissen A.W.R.H., Holub E., van der Hucht J., van den Berg J.A.,  
 RA Steensma H.Y.;  
 RT "Sequence of the open reading frame of the FLO1 gene from  
 RT Saccharomyces cerevisiae.";  
 RL Yeast 9:423-427(1993).  
 RN [4]  
 RP REVIEW.  
 RX MEDLINE=96076625; PubMed=7502576;  
 RA Teunissen A.W., Steensma H.Y.;  
 RT "Review: the dominant flocculation genes of Saccharomyces cerevisiae  
 RT constitute a new subclomeric gene family.";  
 RL Yeast 11:1001-1013(1995).  
 CC - FUNCTION: May be directly involved in the flocculation process.  
 CC - SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor



(Potential).  
 -1- PFM: Extensively O-glycosylated (Probable).  
 -1- SIMILARITY: Belongs to the flocculin family.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL: X78160; CAAS5024.1; -  
 DR EMBL: L28920; AAC09499.1; ALT\_SEQ.  
 DR PIR: S53465; S53465.  
 DR Germonline: L38425; -  
 DR SGD: S0000084; FLO1.  
 DR GO: GO:0009277; C:cell wall (sensu Fungi); IDA.  
 DR GO: GO:0005194; F:cell adhesion molecule activity; IDA.  
 DR GO: GO:000128; P:flocculation; IMP.  
 DR InterPro: IPR001389; Flocculin.  
 DR Pfam: PF00624; Flocculin; 18.  
 KM Glycoprotein; Membrane; Repeat; Cell wall; Signal; GPI-anchor.  
 FT SIGNAL 1 24  
 FT CHAIN 1 25  
 FT PROPEP 2 1537  
 FT DOMAIN 274 1240  
 FT REPEAT 278 1087  
 FT REPEAT 278 322  
 FT REPEAT 323 367  
 FT REPEAT 368 412  
 FT REPEAT 413 457  
 FT REPEAT 458 502  
 FT REPEAT 503 547  
 FT REPEAT 548 592  
 FT REPEAT 593 637  
 FT REPEAT 638 682  
 FT REPEAT 683 727  
 FT REPEAT 728 772  
 FT REPEAT 773 817  
 FT REPEAT 818 862  
 FT REPEAT 863 907  
 FT REPEAT 908 952  
 FT REPEAT 953 997  
 FT REPEAT 998 1042  
 FT REPEAT 1043 1087  
 FT DOMAIN 1118 1157  
 FT REPEAT 1118 1137  
 FT REPEAT 1138 1157  
 FT DOMAIN 1226 1392  
 FT REPEAT 1226 1276  
 FT REPEAT 1291 1341  
 FT REPEAT 1342 1392  
 FT DOMAIN 1408 1434  
 FT REPEAT 1408 1416  
 FT REPEAT 1417 1425  
 FT REPEAT 1426 1434  
 FT REPEAT 1435 1435  
 FT CARBOHYD 187 187  
 FT CARBOHYD 262 262  
 FT CARBOHYD 329 329  
 FT CARBOHYD 374 374  
 FT CARBOHYD 419 419  
 FT CARBOHYD 509 509  
 FT CARBOHYD 554 554  
 FT CARBOHYD 599 599  
 FT CARBOHYD 644 644  
 FT CARBOHYD 689 689  
 FT CARBOHYD 734 734  
 FT CARBOHYD 1114 1114  
 SQ SEQUENCE 1537 AA; 160722 MW; 992650C6B9ABCEA CRC64;

Query Match 4.7%; Score 130.5; DB 1; Length 1537;

Best Local Similarity 20.0%; Pred. No. 1;  
 Matches 119; Conservative 97; Mismatches 236; Indels 143; Gaps 27;  
 -----  
 QY 11 GLVCTGLANVSKRALDLWSLNEATVARTALINMGADGAVSGADSGIVASPESTDNP 70  
 DB 769 GLIST-----TTEPW-TGTFSTSTEMTTVTGNGQPTD--ETVIVIRPTSEG- 814  
 QY 71 DYRYTWTRD--SGLVAKTVLDLFRNGDTSLSITENYI-----SAOAVIGISNP--SGDL 122  
 DB 815 --LVTTTTPMTGTFSTSTEMTTITGNGVPRIDEVIVIRPTSGLISTTTEPTMTGT 872  
 QY 123 SSGAGLGEKPFVNDFTAVTSGWGRPQD-----GPALRATAMIGGQWLLDNGYSTAT 176  
 DB 873 TS-----TSTEMTTITGNGQPTDEIVIVIRPTSGLISTTTEPW--TGTFSTST 922  
 QY 177 D-----IWPVANDLSYVAQYNNQGYDMEVY-----NGS-SFF 211  
 DB 923 EMTVTGTNGVPTDEIVIVIRPTSEGLISTTTEPTGTFTSTSTVTITGNGQPTDE 982  
 QY 212 TIAVQHALVEGSAFATAVGSSCSWCDSQAPETLCYLGQFWTGSFILANPD-----SSRS 267  
 DB 983 TVIVIRPTSEGLISTT-----EPMGTFTSTEMTTVTGNG 1022  
 QY 268 KANTLLGSIHPTDPACDSTFQPCSPRALANHKVVDSPRSYITLNDGLSDSAVAV 327  
 DB 1023 QPTDEIVIVIRPTSEGLVTTT--EPMGTFTSTSTEM-----STVTGNGLPTDEIVIV 1076  
 QY 328 GRYP-----EDTYNGNPMFLCTLAABEQLYDMXQGLSEVTDVSLDFKALYSDA 382  
 DB 1077 VKPTPTAISSSSSSQITSSIRPIITPFPSNGTSSVSSVSSVTSLSFTSS 1136  
 QY 383 ---ATGYSSSSSTVSIV-DAVKTFADGFSIVETHAASNGMSQYKSDG-----431  
 DB 1137 PVISSSVISSSTTSTISPESSKS-----SVIPISSSTSGS--SSSTSSAGSVSSSR 1189  
 QY 432 -EQLSARDLTWSYALLTANRRNSVVPASGSETASVPGTCAATSAIGYSSVTVTSW 490  
 DB 1190 ISSESSKPTTSSSSLPVTSATTS-----QETASLPPATTTKTSSEQTLL--VVTSC 1241  
 QY 491 PSIVATGTT---TTATPTGSGSVT-----STKTTATASKTSTTR 529  
 DB 1242 ESHVCTESISPAIVSATVTVSGVTEYTWCPISYETTKQKGTTEQTETTK 1296  
 -----  
 RESULT 43  
 ID TENA\_PIG STANDARD; PRT; 1746 AA.  
 AC Q29116; P98142;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Tenascin precursor (TN) (Hexabrachion) (Cytocactin) (Neuronectin)  
 DE (GEMM) (J1) (Mitochondrial antigen) (Glioma-associated-extracellular  
 DE matrix antigen) (GP 150-225) (Tenascin-C) (TN-C) (P230).  
 GN TNC OR HXB.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OC NCBI\_TaxID=9823;  
 RN (1)  
 RP SEQUENCE FROM N.A. (ISOFORMS MAJOR; MINOR-1 AND MINOR-2).  
 RC TISSUE=Submaxillary gland;  
 RX MEDLINE=92104189; PubMed=1722152;  
 RA Nishi T., Weinstein J., Gillespie W.M., Paulson J.C.;  
 RT "Complete primary structure of porcine tenascin: detection of  
 RL tenascin transcript in adult submaxillary glands.";  
 RL Eur. J. Biochem. 202:643-648 (1991).  
 RN [2]  
 RP SEQUENCE OF 813-825; 887-917; 998-1011; 1597-1608 AND 1719-1730.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=98158323; PubMed=9498558;  
 RA Wakatsuki S., Ho S.H., Arioka M., Yamasaki M., Kitamoto K.;  
 RT "Isolation and characterization of a 230 kDa protein (p230)



specifically expressed in fetal brains: its involvement in neurite outgrowth from rat cerebral cortex neurons grown on monolayer of astrocytes.";

J. Biochem. 122:1146-1152(1997).

-1 FUNCTION: SAM (substrate-adhesion molecule) that appears to inhibit cell migration. May play a role in supporting the growth of epithelial tumors. Is a ligand for integrins alpha-8/beta-1, alpha-9/beta-1, alpha-v/beta-3 and alpha-v/beta-6.

-1 FUNCTION: Plays a role during early brain development particularly in growth cone guidance. Involved in neurite outgrowth from cortical neurons grown on the monolayer of astrocytes.

-1 SUBUNIT: Hexameric. A homotrimer may be formed in the triple coiled-coil region and may be stabilized by disulfide rings at both ends. Two of each half-hexadachions may be disulfide linked within the central globule.

-1 SUBCELLULAR LOCATION: Secreted; extracellular matrix.

-1 ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=3;  
Comment=isoforms are produced in a tissue-and time-specific manner during development;

Name=Minor-2;  
IsoId=Q29116-1; Sequence=Displayed;

Name=Major;  
IsoId=Q29116-2; Sequence=VSP\_001416;  
Name=Minor-1;  
IsoId=Q29116-3; Sequence=VSD\_001417;

-1 TISSUE SPECIFICITY: Submaxillary glands and brain.

-1 DEVELOPMENTAL STAGE: Predominantly expressed in the embryonic and early postnatal stages. Little or no detection in adult brain.

-1 INDUCTION: By TGF-beta.

-1 SIMILARITY: Contains 15 EGF-like domains.

-1 SIMILARITY: Contains 11 fibronectin type III domains.

-1 SIMILARITY: Contains 1 fibronogen C-terminal domain.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; X61599; CAA43796.1; -.  
PIR; S19694; S19694.  
HSSP; P24821; ITEN.  
InterPro: IPR006209; EGF like.  
DR InterPro: IPR002181; Fibrinogen\_C.  
InterPro: IPR008957; FN\_III-like.  
DR InterPro: IPR003961; FN\_III.  
InterPro: IPR006210; IEFG.  
DR InterPro: IPR002049; Laminin\_EGF.  
Pfam; PF00008; EGF\_12.  
DR Pfam; PF00147; fibrinogen\_C; 1.  
Pfam; PF00041; fn3; 10.  
PRINTS; PR00011; EGFLAMININ.  
DR SMART; SM00181; EGF; 8.  
DR SMART; SM00186; FBG; 1.  
SMART; SMO0060; FN3; 8.  
DR PROSITE; PS00022; EGF\_1; 15.  
DR PROSITE; PS01186; EGF\_2; 14.  
DR PROSITE; PSS0026; EGF\_3; 5.  
KW Glycoprotein; Cell adhesion; Repeat; EGF-like domain; Coiled coil; Extracellular matrix; Alternative splicing; Signal.  
FT SIGNAL 1..22 POTENTIAL.  
FT CHAIN 23..1746 TENASCIN.  
FT DOMAIN 23 185 INVOLVED IN HEXAMER FORMATION.  
FT DOMAIN 118 145 COILED COIL (POTENTIAL).  
FT DOMAIN 174 186 EGF-LIKE 1 (INCOMPLETE).  
FT DOMAIN 187 217 EGF-LIKE 2.  
FT DOMAIN 218 249 EGF-LIKE 3.  
FT DOMAIN 250 280 EGF-LIKE 4.  
FT DOMAIN 281 311 EGF-LIKE 5.  
FT DOMAIN 312 342 EGF-LIKE 6.

FT	DOMAIN	343	373	Egf-Like 7.
FT	DOMAIN	374	404	Egf-Like 8.
FT	DOMAIN	405	435	Egf-Like 9.
FT	DOMAIN	436	466	Egf-Like 10.
FT	DOMAIN	467	497	Egf-Like 11.
FT	DOMAIN	498	528	Egf-Like 12.
FT	DOMAIN	529	559	Egf-Like 13.
FT	DOMAIN	560	589	Egf-Like 14.
FT	DOMAIN	590	620	Egf-Like 15.
FT	DOMAIN	621	710	FIBRONECTIN TYPE-III 1.
FT	DOMAIN	711	801	FIBRONECTIN TYPE-III 2.
FT	DOMAIN	802	891	FIBRONECTIN TYPE-III 3.
FT	DOMAIN	892	983	FIBRONECTIN TYPE-III 4.
FT	DOMAIN	984	1071	FIBRONECTIN TYPE-III 5.
FT	DOMAIN	1072	1162	FIBRONECTIN TYPE-III 6.
FT	DOMAIN	1163	1253	FIBRONECTIN TYPE-III 7.
FT	DOMAIN	1254	1342	FIBRONECTIN TYPE-III 8.
FT	DOMAIN	1343	1430	FIBRONECTIN TYPE-III 9.
FT	DOMAIN	1431	1518	FIBRONECTIN TYPE-III 10.
FT	DOMAIN	1527	1733	FIBRINOGEN C-TERMINAL.
FT	DISULFID	64	64	INTERCHAIN (POTENTIAL).
FT	DISULFID	190	200	BY SIMILARITY.
FT	DISULFID	194	205	BY SIMILARITY.
FT	DISULFID	207	216	BY SIMILARITY.
FT	DISULFID	221	231	BY SIMILARITY.
FT	DISULFID	225	236	BY SIMILARITY.
FT	DISULFID	238	247	BY SIMILARITY.
FT	DISULFID	252	263	BY SIMILARITY.
FT	DISULFID	256	268	BY SIMILARITY.
FT	DISULFID	270	279	BY SIMILARITY.
FT	DISULFID	284	294	BY SIMILARITY.
FT	DISULFID	288	299	BY SIMILARITY.
FT	DISULFID	301	310	BY SIMILARITY.
FT	DISULFID	315	325	BY SIMILARITY.
FT	DISULFID	319	330	BY SIMILARITY.
FT	DISULFID	332	341	BY SIMILARITY.
FT	DISULFID	346	356	BY SIMILARITY.
FT	DISULFID	350	361	BY SIMILARITY.
FT	DISULFID	363	372	BY SIMILARITY.
FT	DISULFID	377	387	BY SIMILARITY.
FT	DISULFID	381	392	BY SIMILARITY.
FT	DISULFID	394	403	BY SIMILARITY.
FT	DISULFID	408	423	BY SIMILARITY.
FT	DISULFID	412	428	BY SIMILARITY.
FT	DISULFID	425	434	BY SIMILARITY.
FT	DISULFID	439	449	BY SIMILARITY.
FT	DISULFID	443	454	BY SIMILARITY.
FT	DISULFID	456	465	BY SIMILARITY.
FT	DISULFID	470	480	BY SIMILARITY.
FT	DISULFID	474	486	BY SIMILARITY.
FT	DISULFID	487	495	BY SIMILARITY.
FT	DISULFID	501	511	BY SIMILARITY.
FT	DISULFID	505	516	BY SIMILARITY.
FT	DISULFID	518	527	BY SIMILARITY.
FT	DISULFID	532	542	BY SIMILARITY.
FT	DISULFID	536	547	BY SIMILARITY.
FT	DISULFID	549	558	BY SIMILARITY.
FT	DISULFID	563	573	BY SIMILARITY.
FT	DISULFID	567	578	BY SIMILARITY.
FT	DISULFID	580	589	BY SIMILARITY.
FT	DISULFID	594	604	BY SIMILARITY.
FT	DISULFID	598	609	BY SIMILARITY.
FT	DISULFID	611	620	BY SIMILARITY.
FT	CARBOHYD	38	38	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	166	166	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	184	184	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	327	327	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	788	788	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	1034	1034	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	1079	1079	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	1121	1121	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	1354	1354	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	VASPSPLIC	1072	1162	Missing (in isoform Major).



FT CARBOHYD 652 652 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 658 658 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 706 AA; 75455 MW; 5E7BF2C77780DC2 CRC64;

Query Match 4.7%; Score 130; DB 1; Length 706;

Best Local Similarity 17.6%; Pred. No. 0.39; Indels 282; Gaps 35;  
 Matches 130; Conservative 111; Mismatches 217;

QY 1 MSFRLSLALSGVCTGLANVI-SKRATLDSML-----SNEATVARTAILNIGADGAW 52  
 DB 1 MQRNLILQASSLI-SGLSLADSSSTTGCGVAPBIIIPCPDDPSILVANA----- 48  
 QY 53 VSGADSGIVVASTNDPNPYFYTRSGVLTKLVDLPFN-----GDTSLSTIENYI 106  
 DB 49 -----SGLSTA--ETD-----WLKRRDAYTKEALHSFLSRATSNPSDLSLTLFPSSN 94  
 QY 107 SAQAVIGSINPSGD---LSSGAGLGEKPFVDETAATGSGMRQRPALRYATAMGF- 162  
 DB 95 SSNVPKIGLACSGGCTYAMLGAGMTIAMDNRIDGANEHGLG-----GLIQSSTYLSGLS 149  
 QY 163 -GQMLD-----NGYTS-----TATDIWPLVR-----NDLSYVAQYV----- 194  
 DB 150 GGMWLTGTALMNMWTSVQEIVDHMSDSIMNITKSIYVPGSGSLVTIIRMSIVGEVQ 209  
 QY 195 --NOTG----DIMEVNGSSFP----- 211  
 DB 210 AKSDAGAGNISLIDMARALSYNFPPLPDAGSALTWSLRDVPFKNGEMPLITVADGR 269  
 QY 212 ----TLIVORALVEGAPATAVAGSSQWCDQAPELCYLQSPWTSFTL---ANFDS 263  
 DB 270 YPGTTVINLNAITLFEIPEFENG---SW---DPSLNAFTDVKYLCTINVTNGKPVAKDQ 320  
 QY 264 SRSKGD-ANTILGSIHTFDEPAACDDSTFQPCSEPRALANKEVDSFRSIYTLNDGLSDS 322  
 DB 321 CVSGYDAGFVIATSAFLNFESLEAST-----STYYKMINSPANKY-VNNISQDD 370  
 QY 323 EAVAVGHPEDTYTNGPM-----FLCTLAALAAQLY----- 353  
 DB 371 DDIAI-----YAANPFQDTEFVDNRNYSIYDADDLFDVGDEGONLPLVLIKKE 422  
 QY 354 ---DALY-----OMDKOG-----SLEVTD-- 369  
 DB 423 RDDDVVFAALISDNTDESMPSGVCMTNTYERQYSKQKGMAFPVPVPTNPLNGLNKP 482  
 QY 370 -----VSLDFPKAL--YSDAATGYSSSSSTYS---SIVDAKTPADGFVSIVET 414  
 DB 483 TFGCDAKNLTDELYIPLVVIYINTGHSFNGSTLKNMYNTERLGMRNGF----- 536  
 QY 415 HAASNGMSQYD-----KSDGEQLSA-----RDLTYSYALALTANNRRN 454  
 DB 537 EAATMGFTDSDNPLGICGAILIRKQESINATLPEPCTKCFADYCWNGTSLSTSAN--- 592  
 QY 455 SVVPASGETSASVPGTCAATSAIG-TYSSVYTWSPSIVANGGTTTATPGSGSVTS 513  
 DB 593 -----PELSGNSYVSGAIAISATDGPITALLGSGTSGNTSNTTSSNVT 645  
 QY 514 TSKTTATASKTSTTTRSGMS 533  
 DB 646 NSNNSNTTLNSNSSSSSIS 665

RESULT 45  
 AMYH YEAST STANDARD; PRT, 1367 AA.

AC P08610; P08068;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Glucoamylase S1/S2 precursor (EC 3.2.1.3) (Glican 1,4-alpha-glucosidase) (1,4-alpha-D-glucan glucosylase).  
 GN STAL OR STAZ OR MAL5 OR YIR019C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 BK Bakayota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 ON NCBI TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=97313266; PubMed=9169870;  
 RA Chutcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,  
 RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,  
 RA Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,  
 RA Moulé S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,  
 RA Skelton J., Smith V., Walsh S., Whitehead S., Barrett B.G.;  
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX";  
 RL Nature 387:84-87(1997).  
 RN [2]  
 RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.  
 RX MEDLINE=87194600; PubMed=3106330;  
 RA Yamashita I., Nakamura M., Fukui S.;  
 RT "gene fusion is a possible mechanism underlying the evolution of  
 RT STAL";  
 RL J. Bacteriol. 169:2142-2149(1987).  
 RN [3]  
 RP SEQUENCE OF 1-31 FROM N.A.  
 RC STRAIN=SPX101-1C;  
 RX MEDLINE=89031230; PubMed=3141213;  
 RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;  
 RT "Similar short elements in the 5' regions of the STZ and SGA genes  
 RT from Saccharomyces cerevisiae";  
 RL FEBS Lett. 239:179-184(1988).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 CC glucose residues successively from non-reducing ends of the chains  
 CC with release of beta-D-glucose.  
 CC -1- SIMILARITY: TO S.POMBE SPEC215.13.  
 CC -1- SIMILARITY: SOME, TO S.POMBE SPEC285.13C.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; Z38061; CA86176.1; -  
 DR EMBL; M16164; AAA35014.1; -  
 DR EMBL; M16165; AAA35015.1; -  
 DR EMBL; X13857; CAA32069.1; -  
 DR PIR; S48478; S48478.  
 DR GenBank; 139731; -  
 DR SGD; S0001458; MUC1.  
 DR GO; GO:0005886; C:plasma membrane, IDA.  
 DR GO; GO:0030447; P:filamentous growth, IDA.  
 DR GO; GO:0007125; P:invasive growth, IMP.  
 DR GO; GO:0007124; P:pseudophal growth, IMP.  
 KW Hydroxylase, Glycosidase, Polysaccharide degradation, Glycoprotein;  
 KW Signal, Multigene family.  
 FT SIGNAL 1 21  
 FT CHAIN 1 21  
 FT DOMAIN 22 1367  
 FT CARBOHYD 817 817  
 FT CARBOHYD 874 874  
 SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CRC64;  
 Query Match 4.7%; Score 129.5; DB 1; Length 1367;  
 Best Local Similarity 20.4%; Pred. No. 1;  
 Matches 111; Conservative 76; Mismatches 211; Indels 147; Gaps 22;  
 QY 29 SGLSNEATVARTAILNIGADGAWSGADSGIVVASTNDPNPYFYTRSGVLTKLTV 88  
 DB 13 SLFNSALGPPTLALVPRGSEGR-----SCNSIVNGCPND-----FNMHMDQONIMQYTL 63  
 QY 89 DLFRNDGTLSTLTENYISQAIVOGISNPSGDLSSGAGLGEKPFVND-----ETAY 140  
 DB 64 DV-----TSVSWQDNTYQITTHVKG-----KENIDLKYLMSLKITGV 101



**THIS PAGE LEFT BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: June 28, 2004, 07:41:12 ; Search time 21 Seconds  
(without alignments)  
2446.013 Million cell updates/sec

Title: US-10-038-723-2  
Perfect score: 2771  
Sequence: 1 MSFRSLALSLGLVCTGLANV.....SKTTATASKSTTTTSGMSL 534

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 100 summaries

Database :  
1: PIR 78:\*  
2: pir1:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2771	100.0	534	2 A29776	glucan 1,4-alpha-g
2	2742	99.0	640	1 ALASGR	glucan 1,4-alpha-g
3	2742	99.0	640	2 A29166	glucan 1,4-alpha-g
4	2620.5	94.6	639	2 JQ0607	glucan 1,4-alpha-g
5	2588.5	93.4	639	2 JTO1346	glucan 1,4-alpha-g
6	1935	69.8	632	2 JTO1346	glucan 1,4-alpha-g
7	1577	56.9	626	2 S36364	glucan 1,4-alpha-g
8	1577	56.9	626	2 T49625	glucan 1,4-alpha-g
9	1497	54.0	493	2 JJC638	glucan 1,4-alpha-g
10	1321	47.7	616	2 T33908	glucan 1,4-alpha-g
11	781.5	28.2	450	2 T39433	glucan 1,4-alpha-g
12	724	26.1	604	2 JPO001	glucan 1,4-alpha-g
13	676.5	24.4	519	1 A54549	glucan 1,4-alpha-g
14	597	21.5	549	1 ALBYG	glucan 1,4-alpha-g
15	522	18.8	778	1 S48474	glucan 1,4-alpha-g
16	521	18.8	767	1 JQ0474	glucan 1,4-alpha-g
17	221.5	8.0	615	2 A64501	glucan 1,4-alpha-g
18	181	6.5	1588	2 A86036	glucan 1,4-alpha-g
19	181	6.5	1588	2 H91188	glucan 1,4-alpha-g
20	172	6.2	622	2 G90250	glucan 1,4-alpha-g
21	170.5	6.2	1275	2 T33369	glucan 1,4-alpha-g
22	166.5	6.0	107	2 B60754	glucan 1,4-alpha-g
23	161	5.8	1063	2 D86731	glucan 1,4-alpha-g
24	160	5.8	888	2 T46726	glucan 1,4-alpha-g
25	159	5.7	1210	2 A25547	glucan 1,4-alpha-g
26	159	5.7	2271	2 F90073	glucan 1,4-alpha-g
27	158	5.7	1283	2 T39174	glucan 1,4-alpha-g
28	157	5.7	1258	2 JQ0188	glucan 1,4-alpha-g
29	156.5	5.6	2232	2 T34434	glucan 1,4-alpha-g

30	155	5.6	1322	2 S07053	ice nucleation pro
31	154.5	5.6	1200	1 SNPSO	ice nucleation pro
32	153	5.5	1331	2 T41144	hypothetical serin
33	151.5	5.5	1034	2 JC2143	ice nucleation act
34	151.5	5.5	1306	2 S25370	MSB2 protein - yea
35	151.5	5.5	1331	2 A48954	mannan endo-1,4-be
36	150	5.4	1609	2 S25345	probable membrane
37	149	5.4	967	2 S66852	hypothetical prote
38	147	5.3	507	2 S64507	probable membrane
39	146	5.3	612	2 C90419	glucan 1,4 alpha g
40	144.5	5.2	614	2 F86719	hypothetical prote
41	144	5.2	948	2 T11678	hypothetical prote
42	142	5.1	1419	2 T30531	agglutinin-like ad
43	141.5	5.1	644	1 T40712	endo-1,4-beta-xyla
44	141.5	5.1	1441	2 B86807	hypothetical prote
45	140.5	5.1	1567	2 S11672	ice nucleation pro
46	140.5	5.1	1567	2 S11672	ice nucleation pro
47	140.5	5.1	1567	2 S11672	ice nucleation pro
48	140	5.1	5188	2 B85547	probable RTX fam1
49	140	5.1	456	2 T38221	hypothetical serin
50	139	5.0	1802	2 S69703	HKR1 protein precu
51	138.5	5.0	402	2 E86185	hypothetical prote
52	138.5	5.0	600	2 S07638	spore coat protein
53	138.5	5.0	1589	2 T13606	hypothetical prote
54	138	5.0	5291	2 F90696	hypothetical prote
55	138	5.0	797	1 VGBEX1	glycoprotein X pre
56	138	5.0	867	2 T45463	membrane glycopro
57	137.5	5.0	625	2 T41603	alpha-amylase - fl
58	137	4.9	556	2 S51892	probable membrane
59	136	4.9	542	2 B47022	chitinase (EC 3.2.
60	136	4.9	866	2 T45462	membrane glycopro
61	136	4.9	995	2 S50358	hypothetical prote
62	135.5	4.9	1180	2 E86719	hypothetical prote
63	135.5	4.9	537	1 JTO150	cellulose 1,4-beta
64	135.5	4.9	537	1 JTO150	cellulose 1,4-beta
65	134	4.8	1350	2 T30294	chitinase (EC 3.2.
66	133.5	4.8	786	2 T16509	large repetitive p
67	133.5	4.8	1145	2 B75625	invariant surface
68	133.5	4.8	694	2 P97279	hypothetical prote
69	132.5	4.8	520	2 S62521	TPR-repeat-contain
70	132.5	4.8	770	2 T22808	hypothetical prote
71	132	4.8	827	2 JC4900	transferred entry
72	131.5	4.7	856	2 T00349	flavocyclase III - As
73	130.5	4.7	1537	2 S53465	flavocyclase prote
74	130.5	4.7	1746	1 S19694	tenascin precursor
75	130	4.7	706	2 S53035	probable lysophosp
76	129.5	4.7	534	2 T33903	serine-rich protei
77	129.5	4.7	1367	1 S48478	glucan 1,4-alpha-g
78	129	4.7	1026	2 A48995	paracrytalline su
79	129	4.7	1032	2 T34433	hypothetical prote
80	129	4.7	1035	2 AD3203	autotransporter pr
81	129	4.7	1073	2 CR7374	S-layer protein R8
82	128.5	4.6	273	2 T44657	protein GP80 [limpo
83	128.5	4.6	1104	2 S59310	probable membrane
84	128	4.6	1169	2 S38181	flucocyclase prote
85	127.5	4.6	776	2 A37035	isoamylase (EC 3.2
86	127.5	4.6	846	2 AD2672	conserved hypochet
87	127.5	4.6	848	2 B97454	hypothetical prote
88	127.5	4.6	881	2 S56032	probable membrane
89	127.5	4.6	1087	2 C84263	transmembrane olig
90	127.5	4.6	3705	2 AD0123	probable autotrans
91	127	4.6	498	1 FLRC	flagellin - Escher
92	127	4.6	540	2 A47022	chitinase (EC 3.2.
93	127	4.6	1016	2 T41720	hypothetical prote
94	127	4.6	1459	2 T32271	hypothetical prote
95	126.5	4.6	463	2 T38444	hypothetical prote
96	126.5	4.6	776	2 S13470	isoamylase (EC 3.2
97	126.5	4.6	1203	2 S27545	pullulanase - Ther
98	126	4.5	389	2 T33340	hypothetical prote
99	126	4.5	609	2 S62518	hypothetical prote
100	126	4.5	973	2 T40778	hypothetical 129.5

ALIGNMENTS

RESULT 1  
A29776  
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) G2 precursor - Aspergillus awamori  
N;Alternate names: glucamylase G2  
C;Species: Aspergillus awamori  
C;Date: 20-Jun-1989 #sequence\_revision 20-Jun-1989 #text\_change 12-Jun-2003  
C;Accession: A93066; A29776  
R;Number, J.H.; Meade, J.H.; Cole, G.; Lawyer, F.C.; McCabe, P.; Schweickart, V.; Tal,  
Mol. Cell. Biol. 4, 2306-2315, 1984  
A;Title: Molecular cloning and characterization of the glucosamylase gene of Aspergillus  
A;Reference number: A93066; MUID:85085934; PMID:6440004  
A;Accession: A93066  
A;Molecule type: DNA  
A;Residues: 1-534 <NN>  
A;Cross-references: GB:K02465; NID:9454405; PID:AA859297.1; PID:gl66506  
R;Number, J.H.; Meade, J.H.; Cole, G.; Lawyer, F.C.; McCabe, P.; Schweickart, V.; Tal,  
submitted to Genbank, February 1985  
A;Reference number: A94514  
A;Contents: annotation; revisions to the DNA sequence and coding regions for G2 form  
C;Comment: The DNA sequence was obtained from Genbank, release 55.0.  
C;Superfamily: glucan 1,4-alpha-glucosidase, yeast type; glucan 1,4-alpha-glucosidase, hydr  
C;Keywords: alternative splicing; extracellular protein; glycoprotein; glycosidase; hyd  
F;23-447/Domain: glucan 1,4-alpha-glucosidase homology <GAG>

Query Match 100.0%; Score 2771; DB 2; Length 534;  
Best Local Similarity 100.0%; Pred. No. 5,9e-179;  
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAILNNGADGAWVSGADSGI 60  
Db 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAILNNGADGAWVSGADSGI 60  
Qy 61 VVASPSTNDPYYTWTNRDSGLVLTVDLFRNGDTSLSLTINNYISQAIVGGINPSG 120  
Db 61 VVASPSTNDPYYTWTNRDSGLVLTVDLFRNGDTSLSLTINNYISQAIVGGINPSG 120  
Qy 121 DLSSGAGLGPKNVDETAATGSGRQPDGPALRATAMIGFGOMLLDNGYSTATDIW 180  
Db 121 DLSSGAGLGPKNVDETAATGSGRQPDGPALRATAMIGFGOMLLDNGYSTATDIW 180  
Qy 181 PLVRNDLSYAAQYNQGYDLMEVNGSSFTTIAVQHRALVEGSAFATAVGSSCSWCDQ 240  
Db 181 PLVRNDLSYAAQYNQGYDLMEVNGSSFTTIAVQHRALVEGSAFATAVGSSCSWCDQ 240  
Qy 241 APEILCYLQSFMTGSLFANPDSRSRGKANTLLGSIHTPDPEACDSTFQPCSPRALA 300  
Db 241 APEILCYLQSFMTGSLFANPDSRSRGKANTLLGSIHTPDPEACDSTFQPCSPRALA 300  
Qy 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPFLLCTTAAAEQLDALYQMD 360  
Db 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPFLLCTTAAAEQLDALYQMD 360  
Qy 361 KQGSLEVTDVSLDFKALYSDAATGTYSSSSSTYSSTIVDAVKTFADGFVSVETHAASNG 420  
Db 361 KQGSLEVTDVSLDFKALYSDAATGTYSSSSSTYSSTIVDAVKTFADGFVSVETHAASNG 420  
Qy 421 SMSROYKSDGEOLASADLTMSYALLTANNRRNSVVPASMGFTSASVPTCAATSAIG 480  
Db 421 SMSROYKSDGEOLASADLTMSYALLTANNRRNSVVPASMGFTSASVPTCAATSAIG 480  
Qy 481 TYSSTVTSWPSIYATGTTTATPTGSGVSTSTKTTATASKTSTTTRSGMSL 534  
Db 481 TYSSTVTSWPSIYATGTTTATPTGSGVSTSTKTTATASKTSTTTRSGMSL 534

RESULT 2  
ALASGR  
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - Aspergillus niger  
N;Alternate names: 1,4-alpha-D-glucan glucosylhydrolase; glucamylase  
C;Species: Aspergillus niger

C;Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 02-Jun-2003  
C;Accession: A90986; A91161; A05287; A22149; A25402  
R;Boel, E.; Hansen, M.T.; Hjort, I.; Hoegh, I.; Fall, N.P.  
EMBO J. 3, 1581-1585, 1984  
A;Title: Two different types of intervening sequences in the glucosamylase gene from Asper  
A;Reference number: A90986; MUID:84261458; PMID:6204865  
A;Accession: A90986  
A;Molecule type: DNA  
A;Residues: 1-640 <BO>  
A;Cross-references: GB:X00712; GB:K02466; NID:92342; PID:CAA25303.1; PID:92343  
A;Note: the authors translated the codon GAT for residue 317 as Asn  
R;Stenson, B.; Larsen, K.; Gunnarsson, A.  
Eur. J. Biochem. 154, 497-502, 1986  
A;Title: Characterization of a glucosamylase G2 from Aspergillus niger.  
A;Reference number: A91161; MUID:8616085; PMID:3081341  
A;Contents: comparison of forms G1 and G2  
A;Accession: A91161  
A;Molecule type: protein  
A;Residues: 25-640 <SV>  
C;Comment: The large molecular form G1 is shown.  
C;Comment: Smaller molecular forms of the enzyme, G2, arise by proteolytic cleavage(s) of  
live towards soluble poly- and oligosaccharides.  
C;Genetics:  
A;Introns: 72/1; 167/3; 200/1; 412/3  
C;Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha-  
C;Keywords: alternative splicing; extracellular protein; glycoprotein; glycosidase; hyd  
F;1-24/Domain: signal sequence #status Predicted <SIG>  
F;23-447/Domain: glucan 1,4-alpha-glucosidase homology <GAG>  
F;25-640/Product: glucan 1,4-alpha-glucosidase G1 #status experimental <GG1>  
F;25-538/Product: glucan 1,4-alpha-glucosidase G2, long form #status experimental <GG2>  
F;25-536/Product: glucan 1,4-alpha-glucosidase G2, short form #status experimental  
F;135,419/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F;465,467,468,477,483,484,492,508,510,513,515,522,525,527,529,530,532/Binding site: carb  
F;476,486,488,496,499,500,502,504,506,512,514,517,518,520,524,526,528,531,534,535/Bit

Query Match 99.0%; Score 2742; DB 1; Length 640;  
Best Local Similarity 99.6%; Pred. No. 6.8e-177;  
Matches 528; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAILNNGADGAWVSGADSGI 60  
Db 1 MSFRSLALSGVCTGLANVSKRATLDSWLSNEATVARTAILNNGADGAWVSGADSGI 60  
Qy 61 VVASPSTNDPYYTWTNRDSGLVLTVDLFRNGDTSLSLTINNYISQAIVGGINPSG 120  
Db 61 VVASPSTNDPYYTWTNRDSGLVLTVDLFRNGDTSLSLTINNYISQAIVGGINPSG 120  
Qy 121 DLSSGAGLGPKNVDETAATGSGRQPDGPALRATAMIGFGOMLLDNGYSTATDIW 180  
Db 121 DLSSGAGLGPKNVDETAATGSGRQPDGPALRATAMIGFGOMLLDNGYSTATDIW 180  
Qy 181 PLVRNDLSYAAQYNQGYDLMEVNGSSFTTIAVQHRALVEGSAFATAVGSSCSWCDQ 240  
Db 181 PLVRNDLSYAAQYNQGYDLMEVNGSSFTTIAVQHRALVEGSAFATAVGSSCSWCDQ 240  
Qy 241 APEILCYLQSFMTGSLFANPDSRSRGKANTLLGSIHTPDPEACDSTFQPCSPRALA 300  
Db 241 APEILCYLQSFMTGSLFANPDSRSRGKANTLLGSIHTPDPEACDSTFQPCSPRALA 300  
Qy 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPFLLCTTAAAEQLDALYQMD 360  
Db 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRYPEDTYNGNPFLLCTTAAAEQLDALYQMD 360  
Qy 361 KQGSLEVTDVSLDFKALYSDAATGTYSSSSSTYSSTIVDAVKTFADGFVSVETHAASNG 420  
Db 361 KQGSLEVTDVSLDFKALYSDAATGTYSSSSSTYSSTIVDAVKTFADGFVSVETHAASNG 420  
Qy 421 SMSROYKSDGEOLASADLTMSYALLTANNRRNSVVPASMGFTSASVPTCAATSAIG 480  
Db 421 SMSROYKSDGEOLASADLTMSYALLTANNRRNSVVPASMGFTSASVPTCAATSAIG 480  
Qy 481 TYSSTVTSWPSIYATGTTTATPTGSGVSTSTKTTATASKTSTTTRSGMSL 530  
Db 481 TYSSTVTSWPSIYATGTTTATPTGSGVSTSTKTTATASKTSTTTRSGMSL 530



Db 481 TYSSVTYVSWPSIVATGGTTTATPTGSGSVTSKTKTATATSKSTSTSS 530

RESULT 3

A29166

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - Aspergillus awamori

N/alternate names: 1,4-alpha-D-glucan glucosylhydrolase precursor; glucanase precursor

C/species: Aspergillus awamori

C/date: 08-Mar-1989 #sequence\_revision 08-Mar-1989 #text\_change 02-Jun-2003

C/accession: A29166

R/Nunberg, J.H.; Meade, J.H.; Cole, G.; Lawyer, F.C.; McCabe, P.; Schweickart, V.; Tal, Mol. Cell. Biol. 4, 2306-2315, 1984

A/title: Molecular cloning and characterization of the glucosylase gene of Aspergillus

A/reference number: A93066; MUID:85085934; PMID:644004

A/accession: A29166

A/molecule type: DNA

A/residues: 1-640 <NN>

A/cross-references: GB:K02465; NID:G454405; PIDN:AA859296.1.; PID:gl66505

C/comment: See also PIR:ALASGR.

C/genetics:

A/introns: 72/1; 167/3; 200/1; 398/3

C/superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha-glucosidase

C/keywords: alternative splicing; extracellular protein; glycoprotein; glycosidase; hydrolase

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-640/Product: glucan 1,4-alpha-glucosidase #status predicted <MAT>

F:23-447/Domain: glucan 1,4-alpha-glucosidase homology <GAG>

Query Match	Similarity	99.0%	Score	2742	DB 2	Length	640
Beat	Local	Similarity	99.6%	Pred.	No. 6.8e-177		
Matches	528	Conservative	1	Mismatches	1	Indels	0
						Gaps	0
Qy	1	MSFRLSLASGLVCTGLANVISKRATLDWSLNEATVARTAILNNGIGADGAWSGASGSI	60				
Db	1	MSFRLSLASGLVCTGLANVISKRATLDWSLNEATVARTAILNNGIGADGAWSGASGSI	60				
Qy	61	VVASPSTDNPDYFYTTWRDSGLVLTLYDLFRNGDTSLSLTENYISAQAIQGISNPSG	120				
Db	61	VVASPSTDNPDYFYTTWRDSGLVLTLYDLFRNGDTSLSLTENYISAQAIQGISNPSG	120				
Qy	121	DLSSGAGLGEKKFNVDETAATYAGSWGRRPORQDPALRATAMIGFGQWLLDNGTSTATDIW	180				
Db	121	DLSSGAGLGEKKFNVDETAATYAGSWGRRPORQDPALRATAMIGFGQWLLDNGTSTATDIW	180				
Qy	181	PLVANDLSYVAQYNQTCYDLMEEVNGSSFETIAVORHALVEGSAFATAVSGSCWCDSQ	240				
Db	181	PLVANDLSYVAQYNQTCYDLMEEVNGSSFETIAVORHALVEGSAFATAVSGSCWCDSQ	240				
Qy	241	APELICYIQSFWTGSPFIANFDSSRSKGNATLLGSIHTFDPBAACDSTFQPCSPRALA	300				
Db	241	APELICYIQSFWTGSPFIANFDSSRSKGNATLLGSIHTFDPBAACDSTFQPCSPRALA	300				
Qy	301	NHKEVNSFRSIYTLNGLSDSEAVAVGRYEDPTYYNGNPMFLCTTLAAEGLYALYQWD	360				
Db	301	NHKEVNSFRSIYTLNGLSDSEAVAVGRYEDPTYYNGNPMFLCTTLAAEGLYALYQWD	360				
Qy	361	KQGLLEVTVDSLDFPKALYSDAATGTYSSSSSSTYSIYDAKTPADGFVSIIVETHAASNG	420				
Db	361	KQGLLEVTVDSLDFPKALYSDAATGTYSSSSSSTYSIYDAKTPADGFVSIIVETHAASNG	420				
Qy	421	SMSEQYDKSDGEOLSDARDLWSYALLTANNRRNSVVPASMGETSASSVPQCAATSAIG	480				
Db	421	SMSEQYDKSDGEOLSDARDLWSYALLTANNRRNSVVPASMGETSASSVPQCAATSAIG	480				
Qy	481	TYSSVTVSWPSIVATGGTTTATPTGSGASTSKTATATSKSTSTRS	530				
Db	481	TYSSVTVSWPSIVATGGTTTATPTGSGASTSKTATATSKSTSTRS	530				

RESULT 4  
J00607  
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - Aspergillus sp.  
NAlternate names: glucocamylase  
iSpecies: Aspergillus sp.

```
C>Date: 31-Mar-1999 #sequence_revision 31-Mar-1999 #ext_change 02-Jun-2003
C/Accession: J00607
R:Shibuya, I.; Gomi, K.; Iimura, Y.; Takahashi, K.; Tamura, G.; Hara, S.
Agric. Biol. Chem. 54, 1905-1914, 1990
A>Title: Molecular cloning of the glucoamylase gene of Aspergillus shiroueami and its ex
A/Reference number: J00607; MUID:91182400; PMID:1368603
A/Accession: J00607
A:Molecule type: DNA
A/Residues: 1-639 <SH1>
A/Experimental source: strain RIB 2504
C/Comment: This enzyme catalyzes the release of glucose from the non-reducing ends of st
C/Genetics:
A:Introns: 72/1; 166/3; 199/1; 411/3
C:Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha-
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-24/Domain: signal sequence #status predicted <Sig>
F:23-446/Domain: glucan 1,4-alpha-glucosidase homology <GAG>
F:25-639/Product: glucoamylase #status predicted <Mat>

Query Match          94.6%; Score 2620.5; DB 2; Length 639;
Best Local Similarity 94.3%; Pred. No. 1e-16;
Matches 500; Conservative 16; Mismatches 13; Indels 1; Gaps 1;
```

QY 1 MSFRLLSLSGVTCGLANVISKRAATLDSMENSEAVATAIINNIGADAWSGADSGI 60

Db 1 MSFRLLSLSGVTCGLASVISKRAATLDSMENSEAVATAIINNIGADAWSGADSGI 60

QY 61 VVASPDSTNDPVFYTWTRDSGLVLKTLVDLFRNGDTSLLSTIENYISAQAIYOGISNPSG 120

Db 61 VVASPDSTNDPVFYTWTRDSGIVLKTLYDLFRNGDPTDLLSTIEHYISSQALLIGVSNPSG 120

QY 121 DLSGAGGEPFENVDENATYTSNGRPGORDGALRTAMIGCGOMLLDMGYSTANDIYW 180

Db 121 DLSG-GGEPFENVDENATYAGSWGRPGORDGALRTAMIGCGOMLLDMGYTSAATEIYW 179

QY 181 PLVRNDLSVYAQYMNQTDGDLWEEVNGSSFFIIAIVOHRLVSGSAFATAVGSSCSMCDSDQ 240

Db 180 PLVRNDLSVYAQYMNQTDGDLWEEVNGSSFFIIAIVOHRLVSGSARATAVGSSCSMCDSDQ 239

QY 241 APEIICYLQSFMTGSFIIANFDPSSRSKQDANTLLGSIHTFDPBACDSDTPOCSPRALA 300

Db 240 APOIICYLQSEFMTGSYIIANFDPSSRSKQDNTLLGISIHTFDPBACDSDTPOCSPRALA 299

QY 301 NHKEVVDSPRSIYTLNDGLSDSEAVAGRPEDTYNGNPMFLCTIAAEQLYDALYQWD 360

Db 300 NHKEVVDSPRSIYTLNDGLSDSEAVAGRPEDTSYNGNPMFLCTIAAEQLYDALYQWD 359

QY 361 KQGSLEIVDVSLDFPKALYSDAATGTYSSTYSISIVAVVTFADGFAFISVETTHAASNG 420

Db 360 KQGSLEIVDVSLDFPKALYSGAATGTYSSTYSISIVAVVTFADGFAFISVETTHAASNG 419

QY 421 SMSBOYDKSDGEQSLARDLTWSYALLTANNRNSVVPASWGETSASSVPGTCAATSATG 480

Db 420 SLSEOPDKSDGELSLARDLTWSYALLTANNRNSVVPASWGETSASSVPGTCAATSASG 479

QY 481 TYSSVTYVSWPSIVATGGTTTATPTGSGSVSTSKTPTASKTSTTTYS 530

Db 480 TYSSVTYVSWPSIVATGGTTTATTTGSGGVTSTSKTPTASKTSTTTYS 529

RESULT 5

UT0479

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - Aspergillus awamori

NlAlternate names: glucocamylase I

CISpecies: Aspergillus awamori

CJDate: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #ext\_change 02-Jun-2003

CJAccession: UT0479

R.Hayashida, S.; Kuroda, K.; Ohta, K.; Kuhara, S.; Fukuda, K.; Sakaki, Y.

AgriC. Biol. Chem. 53, 923-929, 1989

Article: Molecular cloning of the glucocamylase I gene of Aspergillus awamori var. kawachi

A.Reference number: UT0479

A.Accession: UT0479

A.Molecule type: DNA

A;Residues: 1-639 <HAY>  
A;Experimental source: var. kawachi  
C;Genetics:  
A;Gene: GAI  
A;Introns: 72/1; 166/3; 199/1; 411/3  
C;Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha  
C;Keywords: glycosidase; hydrolase; polysaccharide degradation  
F;1-24/Domain: signal sequence #status predicted <SIG>  
F;23-446/Domain: glucan 1,4-alpha-glucosidase homology <GAG>  
F;25-639/Product: glucan 1,4-alpha-glucosidase #status predicted <MAT>  
F;494-538/Region: raw-starch-affinity region

Query Match 93.4%; Score 2588.5; DB 2; Length 639;  
Best Local Similarity 93.8%; Pred. No. 1.5e-166;  
Matches 497; Conservative 15; Mismatches 17; Indels 1; Gaps 1;

```
QY 1 MSFRSLALSGVCTGIA-NVISKRATLDSWLSNEATVARTALINNIGADGAWMSGADSGI 60
DB 1 MSFRSLALSGVCTGIA-NVISKRATLDSWLSNEATVARTALINNIGADGAWMSGADSGI 60
QY 61 VVASPSTDNPDYFYTWRDGLVLTIVDLFRNGDTLSLTENYISAQAIVOGINSNPSC 120
DB 61 VVASPSTDNPDYFYTWRDGLVLTIVDLFRNGDTLSLTENYISAQAIVOGINSNPSC 120
QY 121 DLSSGAGLGEPEKFNVDETAYTGSNGRPQRDGPALATAMIGFGQMLDNGYSTATDIW 180
DB 121 DLSSGAGLGEPEKFNVDETAYTGSNGRPQRDGPALATAMIGFGQMLDNGYSTATDIW 180
QY 121 DLSSGAGLGEPEKFNVDETAYTGSNGRPQRDGPALATAMIGFGQMLDNGYSTATDIW 180
DB 121 DLSSGAGLGEPEKFNVDETAYTGSNGRPQRDGPALATAMIGFGQMLDNGYSTATDIW 180
QY 181 PLVANDLSYVAQVYNNQGYDIMEEVNGSSFFTLAVQHRALVEGSAFATAVGSSCSCWCDQ 240
DB 181 PLVANDLSYVAQVYNNQGYDIMEEVNGSSFFTLAVQHRALVEGSAFATAVGSSCSCWCDQ 240
QY 241 APEILICYQSFWTGSFILLANFDSRSRGKDANTLLGSIHTPEPAACDSTFQPCSPRALA 300
DB 241 APEILICYQSFWTGSFILLANFDSRSRGKDANTLLGSIHTPEPAACDSTFQPCSPRALA 300
QY 240 APEILICYQSFWTGSFILLANFDSRSRGKDANTLLGSIHTPEPAACDSTFQPCSPRALA 299
DB 240 APEILICYQSFWTGSFILLANFDSRSRGKDANTLLGSIHTPEPAACDSTFQPCSPRALA 299
QY 301 NHKEVDSFRSITVLTNDGLSDSEAVAVGRYPEDTYNGNPFELCTLAAEQLYDALYQMD 360
DB 301 NHKEVDSFRSITVLTNDGLSDSEAVAVGRYPEDTYNGNPFELCTLAAEQLYDALYQMD 360
QY 300 NHKEVDSFRSITVLTNDGLSDSEAVAVGRYPEDTYNGNPFELCTLAAEQLYDALYQMD 359
DB 300 NHKEVDSFRSITVLTNDGLSDSEAVAVGRYPEDTYNGNPFELCTLAAEQLYDALYQMD 359
QY 361 KQSGLEVTVDVSLDPEFKALYSDAATGYSSSSSTYSIVDAVKTADPVSIVETHAASNG 420
DB 361 KQSGLEVTVDVSLDPEFKALYSDAATGYSSSSSTYSIVDAVKTADPVSIVETHAASNG 420
QY 360 KQSGLEVTVDVSLDPEFKALYSDAATGYSSSSSTYSIVSAKVTADGVSIVETHAASNG 419
DB 360 KQSGLEVTVDVSLDPEFKALYSDAATGYSSSSSTYSIVSAKVTADGVSIVETHAASNG 419
QY 421 SMSSEQYKSDGEQSLARDLTWSYALTLTANNRRNSVVPASGERTSASSVPCTCAATSAIG 480
DB 421 SMSSEQYKSDGEQSLARDLTWSYALTLTANNRRNSVVPASGERTSASSVPCTCAATSAIG 480
QY 420 SLSSQPKSDGDELSADDLTWSYALTLTANNRRNSVVPASGERTSASSVPCTCAATSAIG 479
DB 420 SLSSQPKSDGDELSADDLTWSYALTLTANNRRNSVVPASGERTSASSVPCTCAATSAIG 479
QY 481 TYSSVTVTSMPSIVATGTTTATPTGSGVSTSTSKTTATASKTSTTSS 530
DB 481 TYSSVTVTSMPSIVATGTTTATPTGSGVSTSTSKTTATASKTSTTSS 530
QY 480 TYSSVTVTSMPSIVATGTTTATPTGSGVSTSTSKTTATASKTSTTSS 529
DB 480 TYSSVTVTSMPSIVATGTTTATPTGSGVSTSTSKTTATASKTSTTSS 529
```

## RESULT 6

QJ01346  
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - *Aspergillus oryzae*  
N;Alternate names: glucoamylase  
C;Species: *Aspergillus oryzae*  
C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 02-Jun-2003  
C;Accession: J01346  
R;Hata, Y.; Tsuchiya, K.; Kitamoto, K.; Gomi, K.; Kumagai, C.; Tamura, G.; Hara, S.  
Gene 108, 145-150, 1991  
A;Title: Nucleotide sequence and expression of the glucoamylase-encoding gene (gluA) frc  
A;Reference number: J01346; MUID:92104497; PMID:1761224  
A;Accession: J01346  
A;Molecule type: DNA  
A;Residues: 1-612 <HAT>  
A;Cross-references: GB:D10698; DDBJ:D01108; NID:g1160312; PIDN:BA01540.1; PID:g1160313  
A;Note: The authors translated the codon TTT for residue 213 as Tyr  
C;Comment: This enzyme hydrolyzes starch to glucose.  
C;Genetics:  
A;Gene: glua  
A;Introns: 75/1; 169/3; 202/1; 414/3

C;Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha  
C;Keywords: glycosidase; hydrolase; polysaccharide degradation  
F;26-449/Domain: glucan 1,4-alpha-glucosidase homology <GAG>

Query Match 69.8%; Score 1935; DB 2; Length 612;  
Best Local Similarity 70.3%; Pred. No. 1.4e-122;  
Matches 367; Conservative 59; Mismatches 90; Indels 6; Gaps 4;

```
QY 1 MSFRSLALSGVCTGIA-NVISKRATLDSWLSNEATVARTALINNIGADGAWMSGADSGI 58
DB 2 VSFSSCLRALALGSSVLAQVPLRQATGDLTWLSTANFSRQALINNIGADGQSAQASGP 61
QY 59 GIYVAPSTDNPDYFYTWRDGLVLTIVDLFRNGDTLSLTENYISAQAIVOGINSNP 118
DB 62 GVYIAPSTDNPDYFYTWRDGLVLTIVDLFRNGDTLSLTENYISAQAIVOGINSNP 121
QY 119 SGDLSSGAGLGEPEKFNVDETAYTGSNGRPQRDGPALATAMIGFGQMLDNGYSTATDI 178
DB 122 SGDLSSGAGLGEPEKFNVDETAYTGSNGRPQRDGPALATAMIGFGQMLDNGYSTATDI 180
QY 179 VWPVANDLSYVAQVYNNQGYDIMEEVNGSSFFTLAVQHRALVEGSAFATAVGSSCSCWCD 238
DB 181 VWPVANDLSYVAQVYNNQGYDIMEEVNGSSFFTLAVQHRALVEGSAFATAVGSSCSCWCD 240
QY 239 SQAPPEILICYQSFWTGSFILLANFDSRSRGKDANTLLGSIHTPEPAACDSTFQPCSPRA 298
DB 241 SQAPPEILICYQSFWTGSFILLANFDSRSRGKDANTLLGSIHTPEPAACDSTFQPCSPRA 300
QY 299 LANHKEVDSFRSITVLTNDGLSDSEAVAVGRYPEDTYNGNPFELCTLAAEQLYDALYQ 358
DB 301 LANHKEVDSFRSITVLTNDGLSDSEAVAVGRYPEDTYNGNPFELCTLAAEQLYDALYQ 360
QY 359 WDKQSGLEVTVDVSLDPEFKALYSDAATGYSSSSSTYSIVDAVKTADPVSIVETHAAS 418
DB 361 WDKQSGLEVTVDVSLDPEFKALYSDAATGYSSSSSTYSIVDAVKTADPVSIVETHAAS 420
QY 419 NSGMSSEQYKSDGEQSLARDLTWSYALTLTANNRRNSVVPASGERTSASSVPCTCAATSA 478
DB 421 TGSMAEQYTKTSDQTSARDLTWSYALTLTANNRRNSVVPASGERTSASSVPCTCAATSA 480
QY 479 IGYSSVTVTSMPSIVATGTTTATPTGSGVSTSTSKTTATASKTSTTSS 518
DB 481 IGYSSVTVTSMPSIVATGTTTATPTGSGVSTSTSKTTATASKTSTTSS 521
```

## RESULT 7

S36364  
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - *Neurospora crassa*  
N;Alternate names: glucoamylase; glycoamylase  
C;Species: *Neurospora crassa*  
C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 02-Jun-2003  
C;Accession: S36364; S13710; S25539  
R;Stone, P.J.; Makoff, A.J.; Parish, J.H.; Radford, A.  
Curr. Genet. 24, 205-211, 1993  
A;Title: Cloning and sequence analysis of the glucoamylase gene of *Neurospora crassa*.  
A;Reference number: S36364; MUID:94037144; PMID:8221928  
A;Accession: S36364  
A;Molecule type: DNA  
A;Residues: 1-626 <SNO>  
A;Cross-references: EMBL:X67291  
R;Koh-Luar, S.I.; Parish, J.H.; Bleasby, A.J.; Pappin, D.J.C.; Ainley, K.; Johansen, F.B.  
Enzyme Microb. Technol. 11, 692-695, 1989  
A;Title: Exported proteins of *Neurospora crassa*: 1-glucoamylase.  
A;Reference number: S13710  
A;Accession: S13710  
A;Molecule type: protein  
A;Residues: 36-60, 'X', 62, 'X', 64-65 <KOH>  
R;Koh-Luar, S.I.; Parish, J.H.; Bleasby, A.J.; Pappin, D.J.C.; Ainley, K.; Johansen, F.B.  
submitted to the Protein Sequence Database, January 1990  
A;Description: Exported proteins of *Neurospora crassa* 1: - glucoamylase.  
A;Reference number: S13711  
A;Accession: S13711  
A;Molecule type: protein

A:Residues: 36-60,'X',62,'X',64-65 <KOW>

C:Genetics:

A:Gene: gla-1

A:Introns: 82/2

C:Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha-glucosidase; hydrolase; polysaccharide degradation

Keyword: glycosylase; signal sequence #status predicted <SIG>

F:1-15/Domain: signal sequence #status predicted <SIG>

F:20-35/Domain: propeptide #status predicted <PRO>

F:33-458/Domain: glucan 1,4-alpha-glucosidase homology <GAG>

F:36-626/Product: glucan 1,4-alpha-glucosidase #status experimental <MAT>

Query Match 56.9%; Score 1577; DB 2; Length 626;

Best Local Similarity 57.6%; Pred. No. 1.9e-98;

Matches 306; Conservative 85; Mismatches 130; Indels 10; Gaps 6;

```

QY 2 SPSRLALSGVLTGTLANYSKATLDSWLSNEATYARFALLNNIGADGAWSGADSGIV 61
DB 13 AFQAVLGLDPLDPEKHKSHDIKR-SVDSYIQETETPAQKLLCNIGASGRASGVAAGV 71
QY 62 VASPTNDPDPYFTWTRDGLVLTVDLFRNG-DTSLSTENTYISQAQIVGINSNSG 120
DB 72 VASPSKSSPDYWTWTRDALVTKLIVDEFTNDYNTLQNTIQAYAAQKLGVSNSG 131
QY 121 DLSGAGLGEPRKFNVDETAYTSGWRPQRDGPALRATAMIGFQOMLNDGYSTATDIW 180
DB 132 SLNKGAGLGEPRKFMVDLQFTGAMGRPQRDGPALRATAMIGYKWLVSNGYADTAKSIW 191
QY 181 PLVRNDLSVVAQYWNQYDLMEEVNGSSFTTAVOHRALVEGSAFATAVGSSCMDSQ 240
DB 192 PLYKNDLAVTAQYWNNTGFDLMEEVNSSFTTIAASHRALVEGSAFAKVSQSCASCDAL 251
QY 241 APEILCYLOSFWTGS-FILANPDSRSRGDANTLGSITFPPEACDDSTPQPCSPRAL 299
DB 252 APQILCFQGSFMSNSGYIISNFVNYRSGKDINSVLTSIHFPDPAACDVNTFPCCDRAL 311
QY 300 ANHKEVVDSPRSIYTLNDGLDSEAVAVGRYPEDTYNGNPMFLCTLAAAEQYDALYQW 359
DB 312 ANHKVVDSPMR-FWGNVSGRTAGKAAVGRYADVYNGNPMFLATLAAAEQYDALYQW 370
QY 360 DKQGSLEVTDLSDPEFKALYSDAATGYSSSSSTYSIVDAVTFADGFSIVETHAASN 419
DB 371 KKGGSITVATSTLAFKDLVPSVSTGYSSSSSTYTAIINAVTYADGFVDIYAQYTPSD 430
QY 420 GSNSEGYDSDGEQLSARDLTMSYALLLTANRRNSVVPASWGETSASVPGCAATSAI 479
DB 431 GSIAEPDQDSGAPLSATHLTMSYASFLSAAARRAGIVPSSWGAASANSIPGSCASATVA 490
QY 480 GTSVSVTVTSWPSIVATGTTT-ATPTGSGS-----VTSTSKTATYASKT 524
DB 491 GSATATATATSPANLTPASTTTPPTGTGCAADHEVLTFNEKVTTSYGQT 541

```

#### RESULT 8

glucan 1,4-alpha-glucosidase (imported) - Neurospora crassa

N:Alternate names: protein B5022.70

C:Species: Neurospora crassa

C:Date: 02-Jun-2000 #sequence, revision 02-Jun-2000 #text\_change 02-Jun-2003

C:Accession: T49625

R:Schulze, U.; Algm, V.; Hohenstein, J.; Brandt, P.; Factmann, B.; Holland, R.; Nykatura,

submitted to the Protein Sequence Database, May 2000

A:Reference number: Z25022

A:Accession: T49625

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-626 <SCH>

A:Cross-references: EMBL:ALJ55932; GSPDB:GN00116; NCSP:B5022.70

A:Experimental source: BAC clone B5022; strain OR74A

C:Genetics:

A:Gene: NCSP:B5022.70

A:Map position: 6

A:Introns: 82/1

C:Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha-

Query Match 56.9%; Score 1577; DB 2; Length 626;

Best Local Similarity 57.6%; Pred. No. 1.9e-98;

Matches 306; Conservative 85; Mismatches 130; Indels 10; Gaps 6;

```

QY 2 SPSRLALSGVLTGTLANYSKATLDSWLSNEATYARFALLNNIGADGAWSGADSGIV 61
DB 13 AFQAVLGLDPLDPEKHKSHDIKR-SVDSYIQETETPAQKLLCNIGASGRASGVAAGV 71
QY 62 VASPTNDPDPYFTWTRDGLVLTVDLFRNG-DTSLSTENTYISQAQIVGINSNSG 120
DB 72 VASPSKSSPDYWTWTRDALVTKLIVDEFTNDYNTLQNTIQAYAAQKLGVSNSG 131
QY 121 DLSGAGLGEPRKFNVDETAYTSGWRPQRDGPALRATAMIGFQOMLNDGYSTATDIW 180
DB 132 SLNKGAGLGEPRKFMVDLQFTGAMGRPQRDGPALRATAMIGYKWLVSNGYADTAKSIW 191
QY 181 PLVRNDLSVVAQYWNQYDLMEEVNGSSFTTAVOHRALVEGSAFATAVGSSCMDSQ 240
DB 192 PLYKNDLAVTAQYWNNTGFDLMEEVNSSFTTIAASHRALVEGSAFAKVSQSCASCDAL 251
QY 241 APEILCYLOSFWTGS-FILANPDSRSRGDANTLGSITFPPEACDDSTPQPCSPRAL 299
DB 252 APQILCFQGSFMSNSGYIISNFVNYRSGKDINSVLTSIHFPDPAACDVNTFPCCDRAL 311
QY 300 ANHKEVVDSPRSIYTLNDGLDSEAVAVGRYPEDTYNGNPMFLCTLAAAEQYDALYQW 359
DB 312 ANHKVVDSPMR-FWGNVSGRTAGKAAVGRYADVYNGNPMFLATLAAAEQYDALYQW 370
QY 360 DKQGSLEVTDLSDPEFKALYSDAATGYSSSSSTYSIVDAVTFADGFSIVETHAASN 419
DB 371 KKGGSITVATSTLAFKDLVPSVSTGYSSSSSTYTAIINAVTYADGFVDIYAQYTPSD 430
QY 420 GSNSEGYDSDGEQLSARDLTMSYALLLTANRRNSVVPASWGETSASVPGCAATSAI 479
DB 431 GSIAEPDQDSGAPLSATHLTMSYASFLSAAARRAGIVPSSWGAASANSIPGSCASATVA 490
QY 480 GTSVSVTVTSWPSIVATGTTT-ATPTGSGS-----VTSTSKTATYASKT 524
DB 491 GSATATATATSPANLTPASTTTPPTGTGCAADHEVLTFNEKVTTSYGQT 541

```

#### RESULT 9

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - Aspergillus oryzae

C:Species: Aspergillus oryzae

C:Date: 05-Dec-1998 #sequence, revision 05-Dec-1998 #text\_change 12-Jun-2003

C:Accession: PC6503; JC6538

R:Hata, Y.; Ishida, H.; Ichikawa, E.; Kawato, A.; Suginami, K.; Imaiya, S.

Gene 207, 127-134, 1998

A:Title: Nucleotide sequence of an alternative glucosylase-encoding gene (glab) express

A:Reference number: JC6538; MUID:98172744; PMID:9511753

A:Accession: PC6503

A:Molecule type: protein

A:Residues: 43-53;71-82;136-145;180-187;327-336;393-404 <HAZ>

A:Experimental source: O-1013

A:Accession: JC6538

A:Molecule type: mRNA

A:Residues: 1-493 <HAT>

A:Cross-references: DDBJ:AB07825

C:Comment: This enzyme is used in industrial processing. It releases glucose from the nor

C:Genetics:

A:Introns: 75/1, 344/2

C:Superfamily: glucan 1,4-alpha-glucosidase, yeast type; glucan 1,4-alpha-glucosidase hom

Keywords: glycoprotein; glycosidase; hydrolase

F:1-26/Domain: signal sequence #status predicted <SIG>

F:26-450/Domain: glucan 1,4-alpha-glucosidase homology <GAG>

F:27-493/Product: glucan 1,4-alpha-glucosidase #status experimental <MAT>

F:139,198,255,369,384,457/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 54.0%; Score 1497; DB 2; Length 493;

Best Local Similarity 56.9%; Pred. No. 3.3e-93;

Matches 277; Conservative 80; Mismatches 128; Indels 2; Gaps 2;

[illegible]

A:introns: 76/1,177/3; 347/2  
C:Superfamily: glucan 1,4-alpha-glucosidase with starch-binding domain; glucan 1,4-alpha-glucosidase; glycoprotein; glycosidase; hydrolase; polysaccharide degradation  
C:Keywords: glycoprotein; glycosidase; hydrolase; polysaccharide degradation  
E:1-25/Domain: signal sequence #status predicted <SIG>  
E:127-455/Domain: glucan 1,4-alpha-glucosidase homology <GAG>  
F:30-616/Product: glucan 1,4-alpha-glucosidase #status experimental <MAT>  
F:199,427/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:1200,514,528,587/Binding site: carbohydrate (Asn) (covalent) #status predicted

```
Query Match          47.7%; Score 1321; DB 2; Length 616;
Best Local Similarity 48.6%; Pred. No. 3,2e-91;
Matches 257; Conservative 88; Mismatches 150; Indels 34; Gaps 6;

14 CTGLANVIS-----KRAITDSWLSNEAVTATAIINIGADGAWYSGADSGIVVA 63
   |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 8 CAGALSLLCSLAIAAPTELKARDLSFISERAILAQALNLSIPDSDAVPGACAGVVA 67

64 SSTNPDPFYFMTWRDGLVLTIVLDFRNGDSTLSTIENYISAQIVOGISNPSGD-L 122
   |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 68 SPSKANPDYFYFMTWRDGLVLTIVLDFRNGDSTLSTIENYISAQIVOGISNPSGD-L 127

123 SSGAGIGEPKENVDETAYTGSWGRPORDGPALRATAMIGFQWLLDNGYSTATDIWPL 182
   |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 128 PDGVLGEGEPKENVDETAYTGSWGRPORDGPALRATAMIGFQWLLDNGYSTATDIWPL 187

183 VANDLSYVAQYNNQCYDIMEEYNGSSPFTIANOHALVGSFAFATVSGSSCSKDSQAP 242
   |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 188 IANDLSYVQYNNQSGFDIMEETVYASSPFTIONOHALVGAQDLADLYTCTGCD-QAP 246

243 ELTCYLOSFWTGSFPIANF--DSSRSKQDANTLGSIHTEPDAACDSTFOPCSPRALA 300
   |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 247 EYLCFLQSGFWNGKIVYISNINANGRTGIDGNSLIGALSTPIDAYCDSPLQPHQSGSLA 306

301 NHKEVVDSPRSIYTLINDGSDSEAVAVGRYPEDTYNGNWFPLCTIAAEQLYDALYQWD 360
   |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 307 NFKVLTDTRNLTYTINAGIPEGQGVAVGRYAEVYNGNENWYLTITAAAEFLYDAVAQWK 366

361 KQGSLEYVDVSLDFFPALISDAATGYSS--SSSTYSIYDAKTPADGVSIVETHAAS 418
   |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 367 AARHVLIVETSIAFFFDIYPELVYREYKSGNANSPQAQIMDAVATAVADSYVALAEKIYPS 426

419 NGSMSQYQKSGDEQSLARDLFTWSYAAALTANNRRNSVVPASGETSASVPGCATSA 478
   |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 427 NGSLSQFQRKRDGTPLSALIDLFTWSYAAFLTMSGRRACQYFSSSGRRALPPTTCSAST 486

479 ICTYSSVTVTSMPSIVATGTTTATPTGSGSVTSSTKTATATASKTSTT 527
   |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 487 PGTY-----TPATAAGAPNTTSCQVSIETFININATT 517
```

RESULT 11  
T39433  
glucosamylase precursor - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 12-Jun-2003  
C:Accession: T39433  
X:1,25/Domain: signal sequence #status predicted <SIG>  
X:127-455/Domain: glucan 1,4-alpha-glucosidase homology <GAG>  
F:30-616/Product: glucan 1,4-alpha-glucosidase #status experimental <MAT>  
F:199,427/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:1200,514,528,587/Binding site: carbohydrate (Asn) (covalent) #status predicted

A:Accession: T39433  
A:Status: preliminary; translated from GB/EMBL/DDDB  
A:Molecule type: DNA  
A:Residues: 1-450 <LN>  
A:Cross-references: EMBL:AL022305; PIDN:CAA18423.1; GSPDB:GN00067; SPDB:SPB14C8.05C  
A:Experimental source: strain 972h-; cosmid c14C8  
A:Genes: SPDB:SPB14C8.05C  
A:Map position: 2  
C:Superfamily: glucan 1,4-alpha-glucosidase, yeast type; glucan 1,4-alpha-glucosidase hom

Query Match 28.2%; Score 781.5; DB 2; Length 450;  
Best Local Similarity 37.6%; Pred. No. 4,4e-45;  
Matches 179; Conservative 68; Mismatches 166; Indels 63; Gaps 12;

QY 7 LALSGLCTG---LANVYSRATLDSWLSNEATAFARAILINIGADGAWMSGADSGIVVA 63  
Db 8 LILGGVISAESLSSPNKRSKEASNDENTDOOKGIAMGHMLNIDSGHMAKDINPGCIIIA 67  
QY 64 SPSTDNDEPYTWTTRDSSLVLKTLLVDLPFRNGDTSLSTIENYSAAOAIVOGISNPISDLS 123  
Db 68 SPSTDSDYYQQWRDSALFTIMTLDRFEEDKGLEPIIVKYEMDEVRLQGVPPSPSDFY 127  
QY 124 SGAGLEGPKFNVDETTAATGSWGRRPQRGPALRAFMIGFGOMLDNGYTSTANDIWPPLV 183  
Db 128 AG-GLGEPRKNRVDDTSYDGWRQPONDSPALRALAFIKYNNYLPEENEKEVEHYEVTWEAV 186  
QY 184 RNDLSYAQYNQGYDLMEEVNSSFFTIAVQHRAIVEGSAPFAVAGSSCSWCDSOAPE 243  
Db 187 LADDYANMHTTEASFDMEEIKDVNHFTILAQGRAMQODGAFAFKRG-----APD 237  
QY 244 -----ILCYLOSFW-TGSFILANFDS--SRSGKDANTLLGSIIHT--FDPEAACD 287  
Db 238 QAAUYORTIEBIDKLGEFMDPGWGVIKGYGRVDRBSGLCSTLIASLYSNEDP----- 291  
QY 288 DSTFPQPSRALANKHEVDSFRSIYTLINDLSDBSAVAVGRYPEDRYN-----GNPWF 342  
Db 292 -----MHLPFLKLQETMTRDYPVNOGSKO-----ANGRYPEDYIDGVKSIGMPWF 339  
QY 343 LCTLAABEQLDALYOWDKOSLEVTVDSLDPFK--ALYSDPAATGYSSSSSTYSIVDA 400  
Db 340 ICTSSAEIITKALAYVDNKGLPELTXYNIHFHKPABFGD-----PYNMSVIRKN 390  
QY 401 VKRPADPFVSIVETHAASNGSMSEFOYDKSDCEQLSADDLTMTSYALLTANNRRRSV 456  
Db 391 MHTYADNFPLAKVAEFQHPNGSMSEOFSRDDHQGABDLTWISSYSLNALTYYRRAT 446

RESULT 12  
JP0001  
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - *Rhizopus oryzae*  
N/Alternate names: 1,4-alpha-D-glucan glucohydrolase; glucoamylase  
C/Species: *Rhizopus oryzae*  
C/Date: 28-Dec-1997 #sequence\_revision 23-Aug-1996 #text\_change 16-Jun-2000  
C/Accession: JP0001  
R/Ashikari, T.; Nakamura, N.; Tanaka, Y.; Ktuchi, N.; Shibano, Y.; Tanaka, T.; Amachi, T.  
Agric. Biol. Chem. 50, 957-964, 1986  
A/Title: *Rhizopus* raw-starch-degrading glucoamylase: its cloning and expression in yeast  
A/Reference number: A90022  
A/Accession: JP0001  
A/Molecule type: DNA  
A/Residues: 1-604 <ASH>  
A/Cross-references: GB:D00049, PIDN:BAA00033.1; PID:g218035  
A/Experimental source: strain SAM0034  
A/Note: there are two errors in the published sequence (personal communication): GCT (11) should be TCT, and the known amino acid sequence of 11 peptides from glucoamylase, including the aminocaproic.  
R/Tanaka, Y.; Ashikari, T.; Nakamura, N.; Ktuchi, N.; Shibano, Y.; Amachi, T.; Yoshizumi, T.  
Agric. Biol. Chem. 50, 965-969, 1980  
A/Title: Comparison of amino acid sequences of three glucoamylases and their structure-F  
A/Reference number: A90023  
A/Contents: annotation; homology; predicted secondary structure  
A/Comment: *Rhizopus glucoamylase* exists in multiple forms, Gluc 1, Gluc 2, and Gluc 3, and degrades raw starch.  
C/Genetics:  
A/Intons: 51/3; 110/3; 129/1; 436/3  
C/Superfamily: *Rhizopus* glucan 1,4-alpha-glucosidase; glucan 1,4-alpha-glucosidase homol  
C/Keywords: glycosidase; hydrolase; polysaccharide degradation  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:36-604/Product: glucan 1,4-alpha-glucosidase (Gluc 1) #status predicted <GL1>  
F:34-138/Domain: glucoamylase starch-binding domain homology <SBD>  
F:156-604/Product: (or 110-604) glucan 1,4-alpha-glucosidase (Gluc 3) #status predicted  
F:159-604/Product: glucan 1,4-alpha-glucosidase (Gluc 2) #status predicted <GL2>  
F:156-592/Domain: glucan 1,4-alpha-glucosidase homology <GAG>  
Query March 26.1k; Score 724; DB 1; Length 604;  
Beet Local Similarity 36.8k; Pred. No. 5e-41;  
Matches 165; Conservative 71; Mismatches 170; Indels 42; Gaps 10;

Qy	25	ATLWSMSENKTVARTATLNNIGADGAWWSGADSGIWAAPSTDNPDNYFYTWTRSDGLVL	84
Db	168	STISSWIKQECISRFMLNRINP-----PGSATGFIAASLSTAGDPDIYANTRDALTS	222
Qy	85	KTLYDLFR--NGDTSLLSTIENYISAQAIYOGISNPGSGLSGAGLGPKEVNDETAYT	141
Db	223	NVIYVEYNTLSTGNKTLINVLKDYVTFSVKTOGSTSVCN-----CLGEPKFPNPDGGYT	276
Qy	142	GSWGRPQDDGALATAMIGFC-QWLLDNGYTSATDIWPLVRNDLSVAAQWYNQGYD	200
Db	277	GAWGRPQDDGAERATFTFLPADSYLTOTKASVYTGTLKPAI PKOLDVYVWVWNSGCFD	336
Qy	201	LMEEVNGSSFFTLIAVQHRALVEGSAFATAVGS--CSWCDSOAPILICLQSEFWTSFIL	258
Db	337	LMEEVNGHFTLWVWRKGLLGGADFAKRGDSTRASTYSSTASTIANKISSFWVSNNW	396
Qy	259	ANFDS-----SRGCKDANTL---LGSIHFTDPEAACDDSTFOQCSPPALANHKEVDS	308
Db	397	IQVSQSVTVGVSKCGLDVSTLLAANLGSV-----DDGFTTPQSEKILAAVAVEDS	447
Qy	309	FRSIYTLNDGGLSDSEAVAVGRYPEDTY-----YNGNPWFCTLLAAEOLYDALYONDKOG	363
Db	448	FASLPIPKULPSYIGNSIGRYPEDTYNGNGNSQGSNPLANTGYALYYRAIKEMINGG	507
Qy	364	SLEVTVDLSLDFFKALYSDAATG-TYSSSSSTYSSISVDVAVKTPADGFVSIIVETHAASNGSM	422
Db	508	GVTWSSISLSPFFKPDSDATSGKTYTGVSDFNNNLAQNIALAADRLSTVQLHHRHNGSL	567
Qy	423	SEQYDKSDGEBQLSARDLTWSYAALLTAN	450
Db	568	AEEFDRTYGLSTGARDLTWASHASILTAS	595

```

RESULT 13
A54549          glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomycopsis fibuligera)
N:Alternate names: 1,4-alpha-D-glucan glucohydrolase; glucamylase
C:Species: Saccharomycopsis fibuligera
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 02-Jun-2003
C:Accession: A54549; S14596
R:Hoshinova, E.; Balanova, J.; Gasperik, J.
FEMS Microbiol. Lett. 67, 103-108, 1991
A:Title: The nucleotide sequence of the glucocamylase gene GlAI from Saccharomycopsis fib
A:Reference number: A54549; MUID:92137640; PMID:1840532
A:Accession: A54549
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-519 <HOS>
A:Experimental source: K2
A>Note: Sequence inconsistent with nucleotide translation
submitted to the EMBL Data Library, March 1991
A:Description: Nucleotide and deduced amino acid sequence of the glucocamylase gene from S.
A:Reference number: S14596
A:Accession: S14596
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-271,'T',273-519 <HO2>
A:Cross-references: EMBL:X58117
C:Genetics:
A:Gene: GlAI
C:Superfamily: glucan 1,4-alpha-glucosidase, Yeast type; glucan 1,4-alpha-glucosidase hom
F,C keywords: glycosidase; hydrolase; polysaccharide degradation
F,I:41-506/Domain: glucan 1,4-alpha-glucosidase homology <GAG>

Query Match      24.4%; Score 676.5; DB 1; Length 519;
Best Local Similarity 36.1%; Pred. NO. 6,4e-38;
Matches 175; Conservative 71; Mismatches 176; Indels 63; Gaps 16;

24 RATDSWISNEATYAKTIILNNIGADGAMVSGADSGIVASSTNDPDTFYTTTRSGTL- 82
|::: : : : : |
|:||||| ||||:| |||||:

```

```

Db      42 RTLETFELDKQXVSLYYLLQNTAYPBGQNDGVPCTVIASPESTSNPDYYYQMTROSAIT 101
Qy      83 ---VLKTLVDLFRNGDSTLSSTIENYISAQAIVOGISNPSG--DLSSAGALGEPKFNVE 137
Db      102 FLTVLSELDE--NNFNTTLAKAVEYINTSYNLTORTSNPGSPFDEKHKGLGEPKFNVDG 159
Qy      138 TAYTGSNGRQORDCPALRATPM-----IGFGMLL---DNGTSTATDITWPLVR 184
Db      160 SATYGAAGRPQNDGPPALRAVAISRYLNDVNSLNGKVLTDSDGIDNPSST-EDYKNIIR 218
Qy      185 NDSIYVAQYNNQGTGYDLMEEVNGSSFFTIAYQHRALVEGSAFATAY-----GSSGSCWD 238
Db      219 PDLEFYVGYDSTGFDLMEENQGRHFTSLVQKAL-----AYAVDIKSPDDGDPASYLD 274
Qy      239 SQAPEILCYLQSFWTG-----SFILANFD-----SSRSKGDANTLLGSHTFPEAACD 288
Db      275 STASTLESISGSDGFVNTDVNHIENPDLIQNSNQGLDSATYIGPLTLHD-IGSSSS 333
Qy      289 STQPCSPRALANKEVVDSPRSITYINDGLSDSEAVAVGRYPEDTY-----YNGNWF 343
Db      334 TFPVDVDEYVLYQSYYLLLENDKORYSVNSAY--SAGAAIGRYPEDVYNGDSSSEGNFWFL 391
Qy      344 CTLLAAEQLYDALYQWMDQSGSLEYT--DVSLEDFKALYSDAT-----GTYSSS 391
Db      392 ATAYAAQVPPKLYVD-AKSNANDITTKINDFPNKITYVDLSTNSGYQSSDSTTIYSGS 450
Qy      392 STYSSIVDAVKTFADEGVSIETHAASNGSMSEQYDKSDGEQLSARDLTWSYALLTANN 451
Db      451 DEFNTVADNLVTFPDSFLQVILDHINDGSLMEDLNRYGTSTAYSLTWSGALLFAIR 510
Qy      452 RKNV 456
Db      511 LRNKC 515

RESULT 14
S48474
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) SGAL - yeast (Saccharomyces cerevisiae)
N:Alternate names: glucoamylase; protein YIL099w
C:Species: Saccharomyces cerevisiae
C>Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 02-Jun-2003
C:Accession: S48474; C26877; E27264
R:Bowman, S.; Churcher, C.
Submitted to the EMBL Data Library, September 1994
A:Reference number: S48474
A:Accession: S48474
A:Molecule type: DNA
A:Residues: 1-549 <BOM>
A:Cross-references: GB:Z47047; EMBL:Z38125; NID:g603997; PID:g763247; GSPDB:GN00009, MIR
R:Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A:Title: Gene fusion is a possible mechanism underlying the evolution of STAI.
A:Reference number: A91831; MUID:87194600; PMID:3106330
A:Accession: C26877
A:Molecule type: DNA
A:Residues: 1-503,507,'W',513-514,'TG',516 <YAM>
A:Cross-references: EMBL:M16166; NID:g172592; PIDN:AAA5042.1; PID:g172593
R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar
A:Reference number: S27281; MUID:89031230; PMID:3141213
A:Accession: S27284
A:Molecule type: DNA
A:Residues: 1-183,'H',185-190 <PAR>
A:Cross-references: EMBL:X13858; NID:g4461; PIDN:CAA32071.1; PID:g4463
C:Genetics:
A:Gene: SGD:SGAL; MIPS:YIL099w
A:Cross-references: SGD:S0001361; MIPS:YIL099w
A:Map position: 9L
C:Function:
A:Description: hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively fr
C:Superfamily: glucan 1,4-alpha-glucosidase, yeast type; glucan 1,4-alpha-glucosidase ho
C:Keywords: glycosidase, hydrolase, polysaccharide degradation; yeast vacuole

```

```

F:77-536/Domain: glucan 1,4-alpha-glucosidase homology <GAG>
Query Match      21.5%; Score 597; DB 1; Length 549;
Best Local Similarity 30.4%; Pred. No. 1,6e-32;
Matches 153; Conservative 89; Mismatches 202; Indels 60; Gaps 14;

Qy      1 MFSRSLALSGLVCTGLANVISKRALDLSNEATYARAILNNTGADGAWSGADSGI 60
Db      56 VQARDVAVLNGTVVVD-SNGAMDSALAEWLQCKKYSIEKTEINTEGPSAVYPS-ISP 113
Qy      61 VVASPETDNPDYFYTWTRDGLVLTVDLFRNGDSTLSSTIENYISAQAIVOGISNPSG 120
Db      114 VIASPQTHDYQYQMTROBALTINSIVS--HSAGPALETLQYLVNSPHLRNN--- 167
Qy      121 DLSSGAG-----LGEKFNVDETAYTGSWGRPPQORDCPALRATPMIGFGMLDNG--- 170
Db      168 TLGAGIGYTDVVALGDPKKNVNDNTAFTEDMGRPQNDGPPALRSIALIKITDYKQSTDL 227
Qy      171 -----YTSTATDITWPLVRNDSLVAQYNNQGTGYDLMEEVNGSSFFTIAYQHRALVEGSA 225
Db      228 GARYPQSTP-DLFDIVRDLRFILDHMNSGFDLMEEVNGMHPFLLVQLSAVDSLS 286
Qy      226 FATAVGSSCSWCD---SQAPEILCYL---QSFWTGSFILANF-----DSSRSKDA 270
Db      287 YFNASERSSPFVEELQTRDISKFLVDPANGFINKY--NYIVGPMIADTLRSGLDI 343
Qy      271 NTLLGSHTFPPEACDDSTFQPCSPRALANKEVVDSPRSITYINDGLSDSEAVAVGRY 330
Db      344 STLLAANTVADAPSA-SHLPEFDINDPAVLNTLHMLHMSIYIPINDSKNATGIALGRY 402
Qy      331 PEDTY-----YNGNPFCTLLAAEQLYDALYQWMDQSGSLEYVDVSLDFKALYSDATG 385
Db      403 PEDVYDGYGEGENPWLATCTASTLYQLYHISQHDLVVPMNDCSNAPWSELVFS 462
Qy      386 TYSS-----SSSTYSSIVDAVKTFADEGVSIETHAASNGSMSEQYDKSDGEQ 433
Db      463 NLITLGNDEGYLLIFERTPAFNQTIQIRPLADSLFKAHAGTQDELSQGNKTYGFM 522
Qy      434 LSARDLTWSYALLTANNRNSV 457
Db      523 QGQHILTWSTYSFWDAYQIQEVL 546

RESULT 15
ALBYG
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - yeast (Saccharomyces cerevisiae)
N:Alternate names: 1,4-alpha-D-glucan glucohydrolase; glucoamylase
C:Species: Saccharomyces cerevisiae
C>Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 05-Sep-1997
A:Accession: A21896; A23470
R:Yamashita, I.; Suzuki, K.; Fukui, S.
J. Bacteriol. 161, 567-573, 1985
A:Title: Nucleotide sequence of the extracellular glucoamylase gene STAI in the yeast Sac
A:Reference number: A21896; MUID:85104778; PMID:3918017
A:Note: S. diastaticus
A:Accession: A21896
A:Molecule type: DNA
A:Residues: 1-778 <YAI>
A:Experimental source: strain 5106-9A; ATCC 60709
R:Yamashita, I.; Suzuki, K.; Sakuro, F.
Agric. Biol. Chem. 50, 475-482, 1986
A:Title: Proteolytic processing of glucoamylase in the yeast Saccharomyces cerevisiae.
A:Reference number: A23470
A:Contents: signal sequence cleavage site
A:Accession: A23470
A:Molecule type: protein
A:Residues: 1-65 <YAZ>
C:Genetics:
A:Gene: STAI
C:Superfamily: yeast glucan 1,4-alpha-glucosidase STAI; glucan 1,4-alpha-glucosidase hom
C:Keywords: extracellular protein; glycoprotein; glycosidase; hydrolase; polysaccharide c
F:33-778/Product: glucan 1,4-alpha-glucosidase #status predicted <MPT>

```



F:345-778/Domain: glucan 1,4-alpha-glucosidase homology <GAG>  
F:46,319,333,425,434,445,524,557,656,661,731,752/Binding site: carbohydrate (Asn) (coval  
F:635/Active site: Asp #status predicted

Query Match 18.8%; Score 522; DB 1; Length 778;  
Best Local Similarity 30.5%; Pred. No. 2,9e-27;  
Matches 139; Conservative 80; Mismatches 177; Indels 60; Gaps 15;

```
QY 1 MSFRLSLASGLCTGLANVISKRATLDSWLSNEATVATATLNNIGADGAWGADSGI 60
DB 324 VQLRDVLMNGTVVYD-SNGAMDSPLLEWLRQKKVSIIRIFENIGPSAVYFS-ILPGV 361
QY 61 VVASPSTDNPDYFYTWTRDSGLVLTVDLFRNGDTSLSTINNYISAQAVOGISNPSG 120
DB 382 VIASPGQTHPDYFYQWIRDSALTNISIVS--HSADPA-ITTLQYLNVSPHLQRTNN--- 435
QY 121 DLSSGAG-----LGEPRNVDEIATYTGSGWRPQRDGPALRATAMIGSQMLDNG--- 170
DB 436 TLGAGIGYNDTVLALDPPKNNVDNTAFTEBPGRPONDGPALRSIALIKIIDIYKOSGTDL 495
QY 171 -----YTSTATDVLWPLVRNDLSVAQYNNQGTGYDLMEEVNGSSFFTIAYQHRALVBSGA 225
DB 496 GAKYFPOSTA-DIFPDIWRMDLRFIIDHNMSSGFDLMEEVNGMFFTLVQLSAVDRSL 554
QY 226 FATAVGSSGSCWD---SOAPEILCYL---QSFMTGSFILANF-----DSSRSQKDA 270
DB 555 YFNASRRSPFVBELAQTRDISKFLVDPRANGFINCKY---NYIVETPMIADTLRSGLDI 611
QY 271 NTLLSGIHFTDPACDDSTFQPCSPRALANKEVDSFRTYTLNDGSLDSEAVAVGRY 330
DB 612 STLAAVTYHADPSA-SHLFPDINDPAVLNTLHMLHMRSIYIPINDSSKNATGIALGRY 670
QY 331 PEDTY-----YNGNPWFCLTAAEQLYDALYQMDKGSLEYVDVSLDFPKALYSDATG 385
DB 671 PEDVYGVGVGEENPWLATCASTTLQYLIRHISEQHDLVVPNNDCSNAPFSELVFS 730
QY 386 TVSS-----SSSTYSSIVDAVKTFADGFV 409
DB 731 NLTLGNDEGYLLFENTPAFNQTIQIRPLADSF 766
```

## RESULT 16

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) STRA2 - yeast (Saccharomyces diastaticus)  
N/Allenate names: glucanase II  
C/Species: Saccharomyces diastaticus  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
R/Accession: J00474; S35895  
R/Lambda: M.G.; Pterotus, I.S.; Sollitt, P.; Marmur, J.  
Gene 100, 95-103, 1991  
A/Title: Primary structure and regulation of a glucanase-encoding gene (STRA2) in Sac  
A/Reference number: J00474; M01D:9127626; PMID:2055484  
A/Molecule type: DNA  
A/Residues: 1-767 <LAM>  
A/Cross-references: GB:M0650; NID:9172733; PIDN:AAA3107.1; PID:9172734  
R/Note: the authors translated the codon CCG for residue 337 as Ala, CCG for residue 364  
submitted to the EMBL Data Library, May 1992  
A/Description: Cloning of a new allelic variant of a Saccharomyces diastaticus glucanase  
A/Reference number: S35895  
A/Molecule type: DNA  
A/Residues: 1-163, 'A', 164-622, 'D', 624-767 <KIM>  
A/Cross-references: EMBL:M0490; NID:9172735; PIDN:AAA20560.1; PID:9172736  
C/Genetics:  
A/Genes: STRA2; DEX1  
C/Superfamily: Yeast glucan 1,4-alpha-glucosidase STRA1; glucan 1,4-alpha-glucosidase hom  
C/Keywords: glycosidase; hydrolase; polysaccharide degradation  
F:334-767/Domain: glucan 1,4-alpha-glucosidase homology <GAG>

Query Match 18.8%; Score 521; DB 1; Length 767;  
Best Local Similarity 30.5%; Pred. No. 3,3e-27;

Matches 139; Conservative 80; Mismatches 177; Indels 60; Gaps 15;

```
QY 1 MSFRLSLASGLCTGLANVISKRATLDSWLSNEATVATATLNNIGADGAWGADSGI 60
DB 313 VQLRDVLMNGTVVYD-SNGAMDSPLLEWLRQKKVSIIRIFENIGPSAVYFS-ILPGV 370
QY 61 VVASPSTDNPDYFYTWTRDSGLVLTVDLFRNGDTSLSTINNYISAQAVOGISNPSG 120
DB 371 VIASPGQTHPDYFYQWIRDSALTNISIVS--HSADPA-ITTLQYLNVSPHLQRTNN--- 424
QY 121 DLSSGAG-----LGEPRNVDEIATYTGSGWRPQRDGPALRATAMIGSQMLDNG--- 170
DB 425 TLGAGIGYNDTVLALDPPKNNVDNTAFTEBPGRPONDGPALRSIALIKIIDIYKOSGTDL 484
QY 171 -----YTSTATDVLWPLVRNDLSVAQYNNQGTGYDLMEEVNGSSFFTIAYQHRALVBSGA 225
DB 485 GAKYFPOSTA-DIFPDIWRMDLRFIIDHNMSSGFDLMEEVNGMFFTLVQLSAVDRSL 543
QY 226 FATAVGSSGSCWD---SOAPEILCYL---QSFMTGSFILANF-----DSSRSQKDA 270
DB 544 YFNASRRSPFVBELAQTRDISKFLVDPRANGFINCKY---NYIVETPMIADTLRSGLDI 600
QY 271 NTLLSGIHFTDPACDDSTFQPCSPRALANKEVDSFRTYTLNDGSLDSEAVAVGRY 330
DB 601 STLAAVTYHADPSA-SHLFPDINDPAVLNTLHMLHMRSIYIPINDSSKNATGIALGRY 659
QY 331 PEDTY-----YNGNPWFCLTAAEQLYDALYQMDKGSLEYVDVSLDFPKALYSDATG 385
DB 660 PEDVYGVGVGEENPWLATCASTTLQYLIRHISEQHDLVVPNNDCSNAPFSELVFS 719
QY 386 TVSS-----SSSTYSSIVDAVKTFADGFV 409
DB 720 NLTLGNDEGYLLFENTPAFNQTIQIRPLADSF 755
```

## RESULT 17

A64501  
glucanase (EC 3.2.1.-) - Methanococcus jannaschii  
C/Species: Methanococcus jannaschii  
C/Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 07-Mar-2003  
R/Accession: A64501  
R/Built: C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirsner, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A/Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Frazer, C.M.; Smith, H.O.; Woese, C  
A/Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
A/Reference number: A64300; M01D:96337999; PMID:8688087  
A/Accession: A64501  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-615 <BUL>  
A/Cross-references: GB:U67601; GB:L77117; NID:92826439; PIDN:AAB99630.1; PID:91592211; T01  
C/Genetics:  
A/Map position: FOR1584527-1586374  
A/Start codon: GTG  
C/Superfamily: Sulfolobus type glucan 1,4 alpha glucosidase  
C/Keywords: glycosidase; hydrolase

Query Match 8.0%; Score 221.5; DB 2; Length 615;  
Best Local Similarity 23.1%; Pred. No. 3,6e-07;  
Matches 101; Conservative 46; Mismatches 125; Indels 165; Gaps 20;

```
QY 59 GIVASPSDNDPDYFYTWTRDSGLVLTVDLFRNGDTSLSTINNYISAQAVOGISNP 118
DB 284 GGIAPSL-HPDYRYWGRD-GSYISIALDLF-----GIRNIPDRFFERMSKIQ----- 331
QY 119 SGLSSGAGIGEKVNDVETATYGSW-----GRPQRGPALRATAM-----IGFQWLL 167
DB 332 -----NAD-----GSLQNYVYNGK-----RLTAITQTDQIGSLTMM 364
QY 168 DNGYTSTANDVWPLVRNDLSVAQYNNQGT-----YDLMEEVNGSSFF 211
```



Db 365 DVHYRLT-----GDRKFEVRYWNTIEKAANYRLVALNFTPCFDMEEHFCVAY 414  
Qy 212 TIAVQHALVEGSAFATV-----GSSCSWCDQAPELLCYLOSFWTGSFILANFD 262  
Db 415 TMGATVAGLKCAVSMSKAVKRDVKOMGKTIEFLKHEVPRF-YLE-----D 461  
Qy 263 SSSSGDANTLLGSIHFDEACDDSTFQPCSPRALANKEVVDSTFRSYTLNDGLSDS 322  
Db 462 EERFASINPL-----DKTIDTSLGLSYFENLID-----VDDEMRMTKT-----A 501  
Qy 323 EAV-----AVGRYPEDTYNGNPMFLCTLA--EOLYDALYQMDKOGSLEVTDV 370  
Db 502 EALEKAFKRVGGRPEIDYREGNPMITTLMLSLYRRLYVLEKDNKA---DI 557  
Qy 371 SLDFPKALYSDAATGYSSSSSTYSSIVDAVKTPADGFSIVETHAASNGSMSEQYDKSD 430  
Db 558 YLQKSKLFFWMVKYSF-----DGLPPEQIHKEI 586  
Qy 431 GEOLSRADLTWSYAL 447  
Db 587 GVPMSAMPLGMSNMF 603

## RESULT 18

A86036  
probable adhesin Z5029 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
C/Species: Escherichia coli  
C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C/Accession: A86036  
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimaiente, E.; Potlomsis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A/Reference number: A85480; MUID:21074935; PMID:11206551  
A/Accession: A86036  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1588 <STO>  
A/Cross-references: GB:A8005174; NID:912518349; PIDN:AAG58749.1; GSPDB:GN00145; UMGF:Z50  
A/Experimental source: strain O157:H7, substrain EDL933  
C/Genetic:  
A/Gene: Z5029

Query Match 6.5%; Score 181; DB 2; Length 1588;  
Best Local Similarity 21.7%; Pred. No. 0.00071;  
Matches 136; Conservative 95; Mismatches 217; Indels 178; Gaps 30;  
Qy 15 TGLANVSKRATLDSWLSNEATVARTALINNIGADG-AWVSGADSGIVASPTDNDYF 73  
Db 576 TNLANTNSNATNTTINISN-----LLETVTNLGEDALKW--DKONGVFTAAHGTEETSKI 628  
Qy 74 YTWTRD-----SGVLKTLVDLFRNGDLSLTSTENYISAQAIYQGISNPSGDL 122  
Db 629 -TNVKOGDLTGTSDAVNGSQLKTTNDAVANTNTNATNTTN-----ISNLETV 677  
Qy 123 SSGAGLGEF--KENVDETAVTGSWGRPQRDGPALRATAMIFGQMLNDNGYSTANDIYW 180  
Db 678 TN--LGEDALKMDKXNGVFTAAHG--NNTASKITN-----ILDGVTATSSDAIN 723  
Qy 181 PLVRNDL-SYVAQW-----NOTG-----YDMEEVNGSSFTTAVQHALVEGSAFA 227  
Db 724 GSQLYDLSNIAITYFGNNAVNTDGVFTGPTYKIGE--TNYNVGDALAAI--NSSFS 777  
Qy 228 TAVGSSCSWCDQAPELLCYLOSFWTGSFI--LANPDSRSKGDA--NTLLGSIHTP--- 280  
Db 778 TSLGDLALW-DATAGKFSAGHGTNGDAVITVDADGEISDSSDAVAVNGSQLHGVSSVVD 836  
Qy 281 ----DEAACDDSTFQPCSPRALANKEVVDSTFRSYTLNDGLSDSEAVAGRYPEDTY 335  
Db 837 ALGGAGVNAVDGTTTAPTYTIANADYDNGDALNAIDTTLDDL-----LMDADAG 887  
Qy 336 YNGNPMFLCTLAAEQYDALYQMDKOGSLEVTDVSLDFPKALYSDAATG----- 365

Db 888 ENG-----AFSAHGDXTASV-TTNVANGAISAASSDAINGSLYTNKYI 933  
Qy 386 -----TYSSSSSTYSSIVDAVTF-----A 405  
Db 934 ADALGGDAEVNADGTTTAPTYTIANAEVNVGDALDLDNALLMDETANGAGAVYASH 993  
Qy 406 DGFVSVIETHAASNGSMSE-QYKSDGEQLSA-----RDLTWSYAL 446  
Db 994 DGRASITTN--VANGSISESTDAVNGSQLNMMIEONTQIINOLAGNTDVTYIENG 1051  
Qy 447 LTNANRNSVVPASWGETSASVPGCAATSAIGTYSVTVTSPSIVATGTTTATPT 506  
Db 1052 AGNINRYTRNDGALPNDASIQGVGAT-----ALG-YNSVAKGDSVAIGGSIQSDVTGI 1105  
Qy 507 GSGSVTSTSKTTPASTSTTRSGM 532  
Db 1106 ALGSSSVSRVIAKGRDSTITENG 1131

## RESULT 19

H91188  
probable adhesin EC64480 [similarity] - Escherichia coli (strain O157:H7, substrain RIMD  
C/Species: Escherichia coli  
C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 02-Nov-2001  
C/Accession: H91188  
R/Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
gasawara, N.; Yaenunga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom  
A/Reference number: A99629; MUID:21156231; PMID:11258796  
A/Accession: H91188  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1588 <HAV>  
A/Cross-references: GB:BA000007; PIDN:BA837903.1; PID:913363955; GSPDB:GN00154  
A/Experimental source: strain O157:H7, substrain RIMD 0509952  
C/Genetic:  
A/Gene: EC64480

Query Match 6.5%; Score 181; DB 2; Length 1588;  
Best Local Similarity 21.7%; Pred. No. 0.00071;  
Matches 136; Conservative 95; Mismatches 217; Indels 178; Gaps 30;  
Qy 15 TGLANVSKRATLDSWLSNEATVARTALINNIGADG-AWVSGADSGIVASPTDNDYF 73  
Db 576 TNLANTNSNATNTTINISN-----LLETVTNLGEDALKW--DKONGVFTAAHGTEETSKI 628  
Qy 74 YTWTRD-----SGVLKTLVDLFRNGDLSLTSTENYISAQAIYQGISNPSGDL 122  
Db 629 -TNVKOGDLTGTSDAVNGSQLKTTNDAVANTNTNATNTTN-----ISNLETV 677  
Qy 123 SSGAGLGEF--KENVDETAVTGSWGRPQRDGPALRATAMIFGQMLNDNGYSTANDIYW 180  
Db 678 TN--LGEDALKMDKXNGVFTAAHG--NNTASKITN-----ILDGVTATSSDAIN 723  
Qy 181 PLVRNDL-SYVAQW-----NOTG-----YDMEEVNGSSFTTAVQHALVEGSAFA 227  
Db 724 GSQLYDLSNIAITYFGNNAVNTDGVFTGPTYKIGE--TNYNVGDALAAI--NSSFS 777  
Qy 228 TAVGSSCSWCDQAPELLCYLOSFWTGSFI--LANPDSRSKGDA--NTLLGSIHTP--- 280  
Db 778 TSLGDLALW-DATAGKFSAGHGTNGDAVITVDADGEISDSSDAVAVNGSQLHGVSSVVD 836  
Qy 281 ----DEAACDDSTFQPCSPRALANKEVVDSTFRSYTLNDGLSDSEAVAGRYPEDTY 335  
Db 837 ALGGAGVNAVDGTTTAPTYTIANADYDNGDALNAIDTTLDDL-----LMDADAG 887  
Qy 336 YNGNPMFLCTLAAEQYDALYQMDKOGSLEVTDVSLDFPKALYSDAATG----- 385  
Db 888 ENG-----AFSAHGDXTASV-TTNVANGAISAASSDAINGSLYTNKYI 933  
Qy 386 -----TYSSSSSTYSSIVDAVTF-----A 405

Db 934 ADALGDADAEVADGTTTAPPTTIANAEYNNVGDALDLDNALLMDETANGAGAYVASH 993  
QY 406 DGFVSIIVETHAASNGSMSE-QYDKSDGEOLSA-----RDLTWVSAAAL 446  
Db 994 DGAASITIN--VANGSISESTDAVNGSQAATNMMEQONTQIINOAGTDTTYIOENG 1051  
QY 447 LTAANNRNSVWPAWGETSASVPGTCAATSAIGTYSVTVTWSPSIIVATGCTTTTATPT 506  
Db 1052 AGINVTYRTNDGGLAFNDASAGVGAT-----AIG-VNSVAKGDSVAIGGGSVDVDTGI 1105  
QY 507 GSGSVTSTKTATTTTASTSTTTTSSGM 532  
Db 1106 ALGSSSVSSRVIAKGRDTSITENG 1131

## RESULT 20

G90250  
glucan 1,4 alpha glucosidase (Glucosylase) [Imported] - Sulfolobus solfataricus  
C/Species: Sulfolobus solfataricus  
C/Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 07-Mar-2003  
C/Accession: G90250  
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.U.; Chan-  
Jong, I.; Ueffing, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F.  
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
submitted to Genbank, April 2001  
A/Description: Sulfolobus solfataricus complete genome.  
A/Reference number: A99139  
A/Accession: G90250  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-622 <KUR>  
A/Cross-references: GB:AE006641; NID:g13814174; PIDN:AAK41262.1; GSPDB:GN00155  
C/Genetics:  
A/Gene: S800990  
C/Superfamily: Sulfolobus type glucan 1,4 alpha glucosidase

Query Match 6.2%; Score 172; DB 2; Length 622;  
Best Local Similarity 21.7%; Pred. No. 0.00078;  
Matches 97; Conservative 56; Mismatches 165; Indels 130; Gaps 19;

QY 37 VARTALNIGADGAVSGADSGIIVASPTNDPDYFTYTRDSGLVLTVLFRNGDT 96  
Db 268 VSLFVIRNHDVNGSIASDSFV-----KTYGDSYQYCPRDAIAYAL-DLAGYKEL 322  
QY 97 SL--LSTIENYIAQAIQVQISNPSGDLSCAGLGEKFNVDATYTGSGRP-QNDGPA 153  
Db 323 ALHGFQFISIANSEGFVYHKNPNTLLAS--MHPWY-----YKGRIRYPIQEDETA 373  
QY 154 LRATAMIGFQMLDNGYTTATDI-----VWPLVRNDLSYVAQYMNQTYDLM 203  
Db 374 LEV-----WAIASHY-EKYEDIDELPLKYPKPAKFMMSFMEGLPKPSFDLME 424  
QY 204 EVNGSFFTIYAVHRLVEGSAFATVAGSSCWCDQAPILCYLQSFMTGSFILANFDS 263  
Db 425 ERYGHIHYTVSYGVALTKGAKLAYDGD-----ILSEDDLS 461  
QY 264 SRGKANTILGS-----IHTPDEP-----AACDDSTFOPCSPRAL-ANHKVDSFRS 311  
Db 462 DTSGLLKGMVLKMTYNGFRRIIDENNODLTVDSSLYAPFFEGVLNANDKIMINTINE 521  
QY 312 I---YTLNDGLSDSEAVAVGRYPEDTY-----NGNPFPLCTLLAAEQLYALQWMDKGS 364  
Db 522 IESRLTVNGGII-----RIENDMYQRKKQPNPIITTL-----W----- 556  
QY 365 LEVTVDSLDFKALYSDAATGYSSSSSTYSIIVDAVKTADGVSIVETHAASNGSMSE 424  
Db 557 -----LSEYVATIND--KNKANEYIKVIRALPTGFLPE 589  
QY 425 QYDKSDGEOLSAADLTYSYAAALTAANR 452  
Db 590 QVDPETFEPTSVPLVWVSHAEFTIAIK 617

## RESULT 21

T33369

hypoetical protein H02P09.3 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 17-Mar-2000

C/Accession: T33369

R;Geisler, C.; Harmon, G.

submitted to the EMBL Data Library, July 1998

A/Description: The sequence of C. elegans cosmid H02P09.

A/Reference number: Z21330

A/Accession: T33369

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1275 &lt;GEI&gt;

A/Cross-references: EMBL:AF075318; PIDN:AA064622.1; GSPDB:GN00028; CESP:H02P09.3

A/Experimental source: strain Bristol N2; clone H02P09

C/Genetics:

A/Gene: CESP:H02P09.3

A/Map position: X

A/Intons: 42/1; 78/3; 106/3; 135/2; 161/2; 224/1

Query Match 6.2%; Score 170.5; DB 2; Length 1275;

Best Local Similarity 22.4%; Pred. No. 0.0027;

Matches 129; Conservative 87; Mismatches 276; Indels 85; Gaps 19;

QY 1 MSFRSLALSGVCTGLANVISKRALDLSWLSNEATVA-RTALNIGADGAVSGADSG 59  
Db 673 VSGSSVTSVSGSTESTSAGASTVSGSTGTSVSDSTISDTNAPGSTESTVTSVS 732  
QY 60 IVVASPTNDPDYFTYTRDS---GLVLTVDL-LFRNGDTSLSTIENTYISAQAIQGI 115  
Db 733 TVSGTSGTSPSTMSASTGNTPGSTESTITDGSYVSGTSGTNNCGSDSSTGI 792  
QY 116 SNESG-DLS--SCAGLGEKFNVDATYTGSGWRPQRDGPALATAMIGFQMLDNGYT 172  
Db 793 STVSGSLSTISGSGSTVSGSSDMTSTGTSSTPSTESTVSGASTMS-----P 842  
QY 173 STATDVLWPLVRNDLSYVAQYMNQ--TYDLMEEVNGSFFTIYAVGH-----R 218  
Db 843 STGSSVETSGSSTVSGSTSSSTTGO---STVESSVSTVSSSTISQSTGTTGE 899  
QY 219 ALVEGSAFATVAGSSCWCDQAPILCYLQSFMTGSFILANFDSRGDANTLLGSIH 278  
Db 900 STVFGSTGSTRATSSSTMSASTGTDTPGSTRSTITTSYVIG--ESTVSGSTITBG-- 955  
QY 279 TPDPEACDSTFOP--CSPRALANHKVDSFRSITLNDGLSDSEAVAVGRYPEDTY 335  
Db 956 -----STISESTWTVGVSTGSTRITGESVSGSTRSTVGESVSGSTESTVSGSTEST- 1009  
QY 336 YNGNPFPLCTLLAAEQLYALQWMDKGSILEVTVDSLDFKALYSDAATGYSSSSSTYS 395  
Db 1010 ----FTVPSTVSGS-----TGSTVTGESTVSGSTASTSSGSTSSSTAGSTVS 1053  
QY 396 SIYDAVKTADGTV-----SIVETHAASNGS--MSRQYKSDGEOLSAADLTYSVA 444  
Db 1054 GSSASTVTSSTGSGSTGEBSTVSGSTVSGSTGSTRITGESVSGSTESTVTSBYSVGS 1113  
QY 445 ALLTANNRR-----NSVPASWGETSASVPGTCAATSAIGTYSVT--VTSPSIIVAT 496  
Db 1114 SVSTGAGNNGSTITGTSVSGSTGSGESTILESSVTSVSGSTITDGSSTRSSVST 1173  
QY 497 GGTITTAATPRG--SGSVTSTKTATTAASKTSTTTBGG 531  
Db 1174 VSASTESTVSGSSASIGSTNTPDSTESTISGTSIG 1210

## RESULT 22

B60754

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) S - creosote fungus (ATCC 20495) (fragments)

C/Species: Anorthotheca resiniae, Hormonoclis resiniae (creosote fungus)

C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 02-Jun-2003

C/Accession: B60754



QY 262 DSSRSGKDNATLLGSIHTFDPPEACDSTFQPCSPALANHKEVDSFRSIYTLANDGLSD 321  
DB 506 ATSSSSDATTSSSSGCT--TSSSDATTSSSDATTSSSEGTATSSSDATTSSGCTAT 563  
QY 322 SEVAVGRPEEDTYVNGNPFLLCTLAAEOLYALQWMDQG--SLEVTVSLDPFPA 377  
DB 564 SSSDVTTSSEGT-----ATSSSDATTSSSEGTSSSDATTSSSEGTAT 609  
QY 378 LVSDATGTYSSTSSSTVSDAVKTFADGFGVIVETHAASNGSMSEQD---KSDGQ 433  
DB 610 TSSDATTSS--SSSEGTSSSDATTSSSD--VTTSSSEGTATSSSDATTSSSEGT 663  
QY 434 LSARDLTWVAALLTANRRNSVVPASWG-----ETSASVPGTCAATSAIGTYSVT 486  
DB 664 TTSSDATTSSSSGCTSSSDATTSSSEGTATSSSDATTSSSEGTSSSDATTSSSE 723  
QY 487 VTSWPSIVATGCTTTTATPFGSSVSTSKTTATKATSTTTRS 530  
DB 724 TATSSDATTSSSEGTSSSDATTSSSEGTATSSSDATTSS 767

RESULT 25  
A25547  
ice nucleation protein - Pseudomonas fluorescens  
C:Species: Pseudomonas fluorescens  
C:Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 26-Aug-1999  
C:Accession: A25547  
R:Warren, G.; Corotico, L.; Wolber, P.  
Nucleic Acids Res. 14, 8047-8060, 1986  
A:Title: Conserved repeats in diverged ice nucleation structural genes from two species  
A:Reference number: A25547; MUID:87040772; PMID:3774551  
A:Accession: A25547  
A:Molecule type: DNA  
A:Residues: 1-1210 <MAX>  
A:Cross-references: GB:X04501; NID:g45507; PIDN:CAA28186.1; PID:g45508  
C:Genetics:  
A:Gene: inaw  
C:Superfamily: ice nucleation protein

Query Match 5.7%; Score 159; DB 2; Length 1210;  
Best Local Similarity 22.2%; Pred. No. 0.015;  
Matches 128; Conservative 72; Mismatches 246; Indels 130; Gaps 25;

QY 8 ALSGLVCTGLANVSKRATLDSWLSNEATVARTAILNNGADGAWMSGADSGIVASPT 67  
DB 165 AVYGSTLTG---ANOSQLAGYSTETAGDSSTLTAAGYSTG--TSGSDSLIAGYST 218  
QY 68 DNPDPFYTWTRDSGLVLTVDLFRNGDNLSTIENYISAQAVOGISNPSGDLSSG-- 125  
DB 219 G-----TAGSDSLIAGYSTOTAGDSSLTA---GYGSTQTAQVGSNLTAGYGSTGA 269  
QY 126 -----AGLGEKENVDETAATGYSWGRPQRDPALRATAMIGFGQMLDNGYTATDI 178  
DB 270 GPDSLSIAGYSTOTAGGESSLTAGYGSTQ-----TAQVGS--LTAAGYGSTG-- 316  
QY 179 VNPVANDLSYVQVKNQGYDLMEEVNGSFTTIAVQHRALVEGSAFATAVSSGSKCD 238  
DB 317 ---AGSDSLIA-----GYGSTQTAQVGSNLTAGYSTQTAQVGSNLTAGYGSTGA 366  
QY 239 SQAPELICYLQSFWTG--SFILANFDSNS--GKDANTLLGSIHTFDPPEACDSTFQ 293  
DB 367 PDSLSIAGYSTOTAGGESSLTAGYGSTQTAQVGSNLTAGYGSTG---AGSDSLIAG 422  
QY 294 C-SPRALANHKEVDSFRSIYTL-----NDGLSDEAVAVGRYPEDTYVNGNPM 341  
DB 423 YGSTQTAGGESSLTAGYGSTQTAQVGSNLTAGYGSTGAGSDSLIAGYGSTQTAGD-- 480  
QY 342 FLCTTAAAEOLYALQWMDQGLETVLVDLDFKALYSDAATGYSSSSSTYSIVDAV 401  
DB 481 --SSLTAG-----YGSTQTAQVGSNLTAGYGSTGAGSDSLIAGYGSS-- 521  
QY 402 KTFADGFGVIVETHAASNGSMSEQDKSDGQLSARDLTWVAALLTANRRNSVVPASW 461

DB 522 -----TGT-AGDSSLTAGYGSTQTAQVGS- DLTAGYGSTGTAGS--DSSLIAG 567  
QY 462 GERTASVPGTCAATSAIGT-----YSSV--TVTSWPSIVATGCTTTATPTG 507  
DB 568 GSTQTAG--GDSLSIAGYGSTQTAQVGSNLTAGYGSTGTAGSDSLIAGYGSTQTA--GG 623  
QY 508 GSVSTSTSKTTATASK-----TSTTRGMSL 534  
DB 624 DSSLTAGYGSTQTAGHGSILTAGYGSTQTAGSSSL 659

RESULT 26  
F90073  
hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: F90073  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguci  
ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kato, C.; Sekimizu, K.; I  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: F90073  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2271 <KID>  
A:Cross-references: GB:BA000018; PID:g13702612; PIDN:BA843752.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA2447

Query Match 5.7%; Score 159; DB 2; Length 2271;  
Best Local Similarity 20.4%; Pred. No. 0.036;  
Matches 116; Conservative 105; Mismatches 263; Indels 84; Gaps 17;

QY 21 ISKRATLDSWLSNEATVARTAILNNGADGAWMSGADSGIVASPTDNPDPFYTWTRDS 80  
DB 748 VTRNSMSDS-VTSGSTQSGSVSTSKADSGASTSTSSIVSTASTSKTSVLSBS 806  
QY 81 GLVLTVDLFRNGDNLSTIENYISAQAVOGISNPSGDLSSGAGLGEKENVDETA 140  
DB 807 VSASKSLSTSESNVS--STSTLVNSQSV---SSSMGGSVSKSTSLSDSISNSTEK 861  
QY 141 TSGWGRPQRDPALRATAMIGFGQMLDNGYTATDIWPL-----VANDLSYVQVKN 195  
DB 862 SSSLSTSTSD--SLRSTSLSDLSMTSTSGSLSKSLSTSTSGSSSTASLSSTSNAI 919  
QY 196 QGYDLMEEVNGSFTTIAVQHRALVEGSAFATAVSSGSKCDSPQPEI-----LC 246  
DB 920 STSTLSSEASTSDSISIS-----NSINSGASTSKSDSOSTSISLSTSDSKMS 970  
QY 247 YLQSF-----WTGSPILA-----NFDSSRSGKDNATLLGSIHTFDPPEACDSTFQ 292  
DB 971 TSESLSDSTSTSGSVSGSLIASQSVSTSDSMSTSEIVSDSISTSGSLASDCKMS 1030  
QY 293 PCSPRALANHKEVDSFRSIYTLNDGLSPSEAVAVGRYPEDTYVNGNPFLLCTLAAEOL 352  
DB 1031 VSSSMSTSGSTSESLSDSOSTSD--SDSKSLT-----STSGSGSTSTSTSTASVTR 1083  
QY 353 YDALYQWMDQGLETVLVDLDFKALYSDAATGYSSSSSTYSIVDAKTFADGFGVIV 412  
DB 1084 SES---QSTSGSVASQSDSMGISTSTSTSDSASASTASSESLISQASISTSGSVST 1140  
QY 413 ETHAASNG-----SMSEQDKSD--GEOLSARDLTWVAALLTANN 451  
DB 1141 TSLSTNSERTSTSVSDSTSLSTSESDSISESTSTSDSISALASSEST--SISLSEMS 1198  
QY 452 RNSVVPASWERTSASVPGTCAATSAIGTYSVTVTSWPSIVATGCTTTA---TPTGS 508  
DB 1199 TSDSESQSAPFLSESLSESTSESVSSSTSESTSLSDSTSESGSTSTSLSNSTSGS 1258  
QY 509 GSV---TSTSKTTATASKSTTTTRGMS 533

Db 1259 ASISTSTISSTSTFKSESVSTSLSMS 1286

## RESULT 27

T39174

hypothetical Serine/threonine repeat containing protein [imported] - fission yeast (Schl  
C/Species: Schlizosacharomyces pombe  
C/Date: 03-Dec-1999 #sequence\_revision 09-Jun-2000 #text\_change 09-Jun-2000  
C/Accession: T50375; T39172; T39173; T39174; T39366  
R/Conor, R.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, February 1998  
A/Reference number: Z21832  
A/Accession: T50375  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1283 <CON>  
A/Cross-references: EMBL:AL021815; PIDN:CAB61533.1; GSPDB:GN00067; SPDB:SPBC8E4.70C  
A/Experimental source: strain 972h-; cosmid c8E4  
A/Accession: T39172  
A/Molecule type: DNA  
A/Residues: 785-1283 <CO2>  
A/Cross-references: EMBL:AL021815; PIDN:CA17000.1  
A/Note: this is an interim translation for a sequence replaced by GenBank/EMBL  
A/Accession: T39173  
A/Molecule type: DNA  
A/Residues: 'ME', '179', 'PLV', '183', 'W', 'KL', '556-761', 'HRGSS' <CO3>  
A/Cross-references: EMBL:AL021815; PIDN:CA17001.1  
A/Note: this is an interim translation for a sequence replaced by GenBank/EMBL  
A/Accession: T39174  
A/Molecule type: DNA  
A/Residues: 1-555, 'S' <CO4>  
A/Cross-references: EMBL:AL021815; PIDN:CA17002.1  
A/Note: this is an interim translation for a sequence replaced by GenBank/EMBL  
R/Mood, V.; Rajandream, M.A.; Barrell, B.G.; Oliver, K.; Harris, D.  
submitted to the EMBL Data Library, March 1999  
A/Reference number: Z21848  
A/Accession: T39366  
A/Status: preliminary; translated from GB/EMBL/DDBJ  
A/Molecule type: DNA  
A/Residues: 1-896, 'G', 898-904, 'I', 906-1283 <MOO>  
A/Cross-references: EMBL:AL035675; PIDN:CA38695.1; GSPDB:GN00067; SPDB:SPBC1289.15  
C/Experimental source: strain 972h-; cosmid c1289  
C/Genetics:  
A/Gene: SPDB:SPBC1289.15; SPDB:SPAC8E4.07C  
A/Map position: 2  
C/Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

Query Match 5.7%; Score 158; DB 2; Length 1283;  
Best Local Similarity 21.0%; Pred. No. 0.019;  
Matches 118; Conservative 76; Mismatches 244; Indels 124; Gaps 24;

QY 5 SLALSLGLVCTGLANVISKRAFLTDSWLSNEA-TVARAILNNTGADAM-VSGADSGIV- 61  
DB 180 SLIPHSITSLTSLVSVINDTASLSKTTSPAGITTEIVSGVSTTFPASGTTSTVE 239  
QY 62 VASPSITNDPVPFYWTTRDGLVLTVDLPFRNGDTSLSTIENYISAOAIVOGISNSGD 121  
DB 240 VVEPT-----AGITTEIV-----SGSVGTSTFPANGTISGIVEVEPTAG- 281  
QY 122 LSSGAGLGEPRKPNVDETAIVYSGWGRPQRDPALRATAMIGFGOMLDNGYTSATDIWVP 181  
DB 282 -----TITFIVSGS-----VGYSFPFANGTISGIVEVEPT 313  
QY 182 LVANDLSVYAQVYKQNTQTYDLMEFVNGSSFTIAYQH-----RALVEGSAFAT---AV 230  
DB 314 TACITTEIVTS--GSVGYTSTFPFANGTISGIVEVEPTAGTVEITIVISGVSIGYSTPEAS 371  
QY 231 GSSCSWCDSDQAPETILCYOSFWTGS-FLANFDSRSRGKANDTLGSIHTFDEPAACDS 269  
DB 372 GTTSGTVEVEPTAGTITTEIVSGSKAFSTTFPAN-----GTTSGTVEVEPTAGTITK 425  
QY 290 TFOPCSPRALANHKEVVDVFRSITYTLNDGLSDSEAV--AVGRYPEDTYNGNFWPLCTL 346

Db 426 TIVSGS-----KTFTSTFPANGTISGIVEVEPTAGTITE-TIVSGSVGYSTTF 473  
QY 347 AAAEQLYDALYQMDKQGSLEVTVDVSLDFPKALYSDAATGYSSSSSTYSIYAVAKTFAD 406  
DB 474 PA-----NOTSGTVEVEPT-----AGITTEIVSGSKTFSTFPASGT-IS 515  
QY 407 GFSIVTEHAASNGNSEQY---DKSDGEQLSARDLTWGYALL--TANNRNSVVPASW 461  
DB 516 GTVEVEPTA---GITTEIVSGSKAFSTFPANGTISGIVEVEPTAGTITEIVSGSV 572  
QY 462 GFTSASVGTCAATGTAIGYSSVTY-----SWPSIVATGTTTATPTGSG 509  
DB 573 GYTSFPASGTTSGTVEVEPTAGTVEITIVSGSVGYSTFPASGTTSGTVEVEPT-AG 631  
QY 510 SVTSTKTATATASKTSTTTRSG 531  
DB 632 TWTETI-VSGSVGYSTTFPASG 652

## RESULT 28

J00188

ice nucleation protein - Erwinia herbicola

C/Species: Erwinia herbicola  
C/Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 26-Aug-1999  
C/Accession: J00188; S35273  
R/Marzen, G.; Corotico, L.  
Gene 85, 239-242, 1989

A/Title: The consensus sequence of ice nucleation proteins from Erwinia herbicola, pseud  
A/Reference number: J00188; MUID:90152370; PMID:2515997  
A/Accession: J00188  
A/Molecule type: DNA  
A/Residues: 1-1258 <WAR>  
A/Cross-references: EMBL:M26382; NID:9148419; PIDN:AAA24823.1; PID:9148420  
A/Experimental source: strain MI  
R/Gurian-Sherman, D.; Lindow, S.E.; Panopoulos, N.J.  
Mol. Microbiol. 9, 383-391, 1993  
A/Title: Isolation and characterization of hydroxylamine-induced mutations in the Erwin  
A/Reference number: S35273; MUID:94018633; PMID:8412688  
A/Accession: S35273  
A/Status: preliminary  
A/Molecule type: protein  
A/Residues: 387-536 <GUR>  
C/Genetics:  
A/Gene: ICE  
C/Superfamily: ice nucleation protein

Query Match 5.7%; Score 157; DB 2; Length 1258;  
Best Local Similarity 22.1%; Pred. No. 0.021;  
Matches 132; Conservative 63; Mismatches 239; Indels 162; Gaps 25;

QY 47 GADAMVSGADSGIVVSPSTNDPVPFYWTTRDGLVLTVDLPFRN-----GPTSLL 99  
DB 299 GADSLIAGYGSQTQAGEBSTQTAGYSTQTAQK-----SLTLAGYSGTGAAGDS-- 350  
QY 100 STIENYISAQAIYQGISNPSG-----DLSSG-----AGLGEPRKN 134  
DB 351 SLIAGYGSQTQAGEBSSLTAGYSGTQTAQKSDLTLAGYSGTGAAGDSLLIAGYSGTQTA 410  
QY 135 VDETAATGSKWRPQRDPALRATAMIGFGOMLDNGYTSATDIWVPVLRANDLSYAQVW 194  
DB 411 GRESTQTAGYSGTQ-----TAQKGS-----LTAAGYSGT-----AGDSSLIA-- 450  
QY 195 NQTYGLMEFVNGSSFTIAYQHRAVLEGSAFATAYGSSCSWCDSDQAPETILCY--IQSF 252  
DB 451 ---GYSTQAGEBSSLTAGYSGTQTAQKSDLTLAGYSGTSTAGYSSSLIAGYSGTQTAG 507  
QY 253 TGSFIILNFDSSRSRGDANTLL---GSIHTFDEPAACDSTFOPCSPRALANHKEVVD- 308  
DB 508 YGSTLTLAGYSGTQTAQNESDLITGYSGTST---AANSSLIAGYSGTQTAASVNSVLTAG 563  
QY 309 FRSITYLNDGL-----SDSEAVA-VGRYPEDTYNGNFWPLCTLAAAEQ-- 351

Db 564 YGSGTQARBSGLTAGYGTAGSDSSIIAGYGTQTAHSYHSLTAGYGTQTAREQSV 623  
QY 352 -----LYDALYQWMDKOSLEVTVDVSLDFPKALYS 380  
Db 624 LTTGCGSTAGADSSLIAGYGTQTAAGNSILTAGY-----GSTQTAQEGSLTAGYGS 678  
QY 381 DAATGYSSSSSYSSIVDAVKTADGVSIVETHAASNGSMEOYDKSDG-----431  
Db 679 TSTAGDSSLIAGYGS-----TQTAGYNSILTAGYGTQTAQEGSDLTSGYGTSTAGA 732  
QY 432 -EQLSA---RDLTWSYALTLTA-----NNRRNSVVPASNGEFSASSVPETCAATSAIG 481  
Db 733 DSSLINGYGTQTAASHSSLTAGYGTQTAAREQSVLTGYSSTAG-----ADSSLIAG 787  
QY 482 YSSVTVTSWPSIVATG-GTTTAA-----PPTGSGSVTSTKTTATATASKTTTTRSG 531  
Db 788 YGSGTQTAHSHILTAGYGTQTAQERSDLTTGYSSTAGADSSLIAGYGTQTAG 843

RESULT 29  
T34434  
hypothetical protein K06A9.1a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T34434  
R:Geisel, C.; Gattung, S.  
Submitted to the EMBL Data Library, December 1996  
A:Description: The sequence of C. elegans cosmid K06A9.  
A:Reference number: Z21525  
A:Accession: T34434  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2232 <GEI>  
A:Cross-references: EMBL:U80846; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a  
A:Experimental source: strain Bristol N2; clone K06A9  
A:Gene: CESP:K06A9.1a  
A:Map position: X  
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/

Query Match 5.6%; Score 156.5; DB 2; Length 2232;  
Best Local Similarity 23.4%; Pred. No. 0.051;  
Matches 113; Conservative 70; Mismatches 228; Indels 71; Gaps 18;

QY 80 SGLVLTLDLFRNGDTSLLSTIENTISQAIVQGISNPSGDISGAGLGEPEKFNDETA 139  
Db 1440 SGGTQHTTMSKASGSGT---SPSTNSQTGSTMWSSSTSGVSTSSASTQPMSTSGS 1496  
QY 140 YTGSMGRPORDGFALATAMIGFQWLLDNGYSTATDIWPLVRNDLSVAQYNNQTCY 199  
Db 1497 SAGSTYASSTASPAASSTAPSSGT-----MSTSSGTYGTTI-SESSTTASASQTG- 1548  
QY 200 DLWEVNGSSFTTIAVQ-----HRLVGSAPATAVAGSSCSW--CDQAPETLCYL 248  
Db 1549 ---STTMGSSSTSGVSTSSASTQPMSTSGSAGSTYASSTAGIVSTVPSSTGM 1605  
QY 249 QSFMTSSFLIANDSRSGKDANTLGSHTTPEPAACDDSTFQPCSPRALANHKEVDS 308  
Db 1606 GSTSSSGT-VGSTISESTTASASQTGSTMWSSSTSGVSTSSASTQPMSTSGSSA 1664  
QY 309 FRSIYTLNDGLDSEAVAVGRYPEDTYVNGNP---WFLCTIAAEQLYDLLYQWMD--Q 362  
Db 1665 GSTVASTTGL-----VSTSTVPSSTGTMSTSGVSGTSSSTRASASSTGSTM 1719  
QY 363 GSLEVTVDVSLDFPKALYSDAATG---TVSSSSSTVSSIVDAVKTADGVSIVETHAAS 418  
Db 1720 GSSSTSGVST-----SSASSGQPMSTSGSSAGSTVVSSTAPAS-----STAPSS 1767  
QY 419 NGSMEOYDKSDGQSLSARDLTWSYALTLANNRRNSVVPASG-----ETGASSVPTG 472  
Db 1768 TGTMSSTSGTGVSTMSQ-----SSTPAASTTSHGTGTVTLGSSSTSSNMQSTSGSSVGS 1822  
QY 473 CAATSAIGTVSSVTVTSWPSIVATGTTTATPTGSGSVTSTKTATA-SKTSITTRSG 531

Db 1823 TVASSSTAGLVSTSTV---PSSGTGWTGSSGT---VGSSTISESTTASASQTGSTVTWG 1876  
QY 532 MS 533  
Db 1877 SS 1878

RESULT 30  
S07053  
ice nucleation protein inaA - Erwinia ananas  
C:Species: Erwinia ananas  
C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 26-Aug-1999  
C:Accession: S07053  
R:Abbe, K.; Matabe, S.; Emori, Y.; Watanabe, M.; Arai, S.  
FEBS Lett. 258, 297-300, 1989  
A:Title: An ice nucleation active gene of Erwinia ananas. Sequence similarity to those of  
A:Reference number: S07053; M0ID:90092494; PMID:2599095  
A:Accession: S07053  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-1322 <ABE>  
A:Cross-references: GB:X17316; NID:9296095; PIDN:CAA35194.1; PID:9296096  
C:Superfamily: Ice nucleation protein

Query Match 5.6%; Score 155; DB 2; Length 1322;  
Best Local Similarity 22.0%; Pred. No. 0.031;  
Matches 125; Conservative 69; Mismatches 220; Indels 154; Gaps 27;

QY 47 GADGAWVGADSGIIVASPDNDPFIYTWTRDSGLVLTLDLFRN-----GDTSL 99  
Db 315 GADSSLIAGYGTQTAARESTQTAGYGTQTAQKG-----SDLTAGYGTGTAQDGS-- 366  
QY 100 STIENTISQAIVQGISNPSG-----DLSSG-----AGLGEPEKFN 134  
Db 367 SLIAGYGTQTAAREBSLTAGYGTQTAQKGSDLTAGYGTGTSAGDSSLIAGYGTQTA 426  
QY 135 VDEATYTGSMGRPORDGFALATAMIGFQWLLDNGYSTATDIWPLVRNDLSVAQY 194  
Db 427 GRESTTQTAGYGTQ-----TQKQSD---LTAGYGTGT-----AGDSSLIAGY 469  
QY 195 N---QTYDLWEVNGSSFTTIAVQHRLVGSAPATAVAGSSCSWCDQAPETLCYLOS 250  
Db 470 STQTAQKGSGL---TAGYGTSTAGYESSLIAG-----YGSF-----QT 505  
QY 251 FMTGSEFLANFPSSNGKQANTLL---GSHTFPEPAACDDSTFQPCSPRALANHKEV 307  
Db 506 AGYSTLTAGYGTQTAQWESDLITGYGTST---AGANSSLIAGYGTQTAASVNSVLT 561  
QY 308 S-FRSIYTLND-----GLSDSEAVAVGRYPEDTYVNGNPWFLC-----TLAA 350  
Db 562 AGYGTQTAAREBSGLTAGYGTGTAQSDSSITAGYGTSTAGADSSLIAGYGTQTAG 621  
QY 351 QLYDALYQWMDKOSLEVTVDVSLDFPKALYSDAATGYSSSSSYSSIVDAVKTADG 410  
Db 622 SILTAGY-----GSTQTAAREBSGLTAGYGTSTAGADSSLIAGYGS-----TQ 670  
QY 411 IV-----ETHAASNGS-MSEQYDKSD-----GEQLSA---RDLTWSYALTLTA 449  
Db 671 ILTAGYGTQTAQEGSDLTAGYGTSTAGADSSLIAGYGTQTAASHSSLTAGYGTQTA 730  
QY 450 NNRRNSVVPASNGEFSASSVPETCAATSAIGTVSSVTVTSWPSIVATG-GTTTAA----- 503  
Db 731 --QEOSVLTTGYSSTAG-----ADSSLIAGYGTQTAAGNSILTAGYGTQTAQERS 783  
QY 504 TPTGSGSVTSTKTTATATASKTTTTRSG 531  
Db 784 LTTGCGSTAGADSSLIAGYGTQTAG 811

RESULT 31  
SNP50  
ice nucleation protein [validated] - Pseudomonas syringae



N,Alternate names: Snomax (TM)  
C:Species: Pseudomonas syringae  
C>Date: 31-Mar-1988 #sequence\_revision 21-Jan-1997 #text\_change 15-Sep-2000  
C:Accession: A24405  
R:Green, R.L.; Warren, G.J.  
Nature 317, 645-648, 1985.  
A>Title: Physical and functional repetition in a bacterial ice nucleation gene.  
A:Reference number: A24405  
A:Molecule type: DNA  
A:Residues: 1-1200 <GRE>  
A:Cross-references: EMBL:X03035; NID:g45828; PIDN:CAA26837.1; PID:g45829  
R:Kajava, A.V.; Lindow, S.E.  
submitted to the Brookhaven Protein Data Bank, June 1993  
A:Reference number: A51242; PDB:1YNA  
A:Contents: annotation; theoretical model, residues 490-535  
R:Kajava, A.V.; Lindow, S.E.  
J. Mol. Biol. 232, 709-717, 1993  
A>Title: A model of the three-dimensional structure of ice nucleation proteins.  
A:Reference number: A58442; MUID:93360260; PMID:8355267  
A:Contents: annotation; theoretical model  
C:Comment: Found on the outer membrane of the bacteria, this protein stimulates ice form  
C:Superfamily: ice nucleation protein  
C:Keywords: tandem repeat  
F:208-1151;Region: 8-residue repeats (A-G-Y-G-S-T-L-T)

Query Match 5.6%; Score 154.5; DB 1; Length 1200;  
Best Local Similarity 22.9%; Pred. No. 0.029;  
Matches 144; Conservative 69; Mismatches 254; Indels 163; Gaps 29;

QY 9 LSGLVCTGLANVISKRATLDLSWLSNEATVARTALINNIGADAWSGADSGIVVASFSTD 68  
DB 207 IAGGSGTGTAG-----SDSWL-----VAGYSGTGTAGDSDALTAGYSGTGTARBSGNL 254  
QY 69 NPYFYTWI--RDSGLVLTGLVDFRNGDPSLSTIENYISQAIVOGISNPGDLSGG- 125  
DB 255 TAGGSGTGTAGSDSLIAGYSGTGTSGDSSLTA--GYSGTGTAGGSGMLTAGYSGTGT 311  
QY 126 -----AGLGPKFNVDETAAYGSGWRPQ--RDGPALRA-----TAMGFG 163  
DB 312 AGSDSLIAGYSGTGTSGDSSLTAAGYSGTGTARBSGNLTAAGYSGTGTAGVDSLIAGY 371  
QY 164 QWL-----LDNGYSTATDIPWLVNNDLSYVAQYWNQGYD 200  
DB 372 STGTSGSDSLTAGYSGTGTARBSGNLTAAGYSGTGT-----AGDSSLIA-----GYG 419  
QY 201 LMEVNSGSGFFITAV-OHRLVBSAATAVGSS-CWCDSQAPELLCTYQSFWTGSFIL 258  
DB 420 STGTSGSDSLTAGYSGTGTARBSGNLTAAGYSGTGTAGVDSLIAGYSGTGTSGDALT 479  
QY 259 ANPDSRSGDANTL--LGSIHFPDEACDDSTFQPC-SPRALANHKEVDSFRSIYT 314  
DB 480 AGGSGTGTAGGSGMLTAGYSGTGT-----AGADSLIAGYSGTGTSGSESSLTAGYSGTGT 535  
QY 315 LNDGL-----SDSEAVAVGRYPEDITYNCPWFLCTLAALPOLYDALY--Q 358  
DB 536 ARGGSTLTAGYSGTGTAGADSSLIA--GYSGTGT-----SGSESSLTAGYSGTGT 582  
QY 359 WDQGSLEIVDVLDFPKALYSDAATGTYSSSSTYSIYDA--VKTFADGPGSIYVTHA 416  
DB 583 TAQGGSVLTGSGYGTGTAGAASNLTTGYSGTGTAGHSSFTIAGYSGTGTAGHKSIL--T 639  
QY 417 ASNGSMSEQYDKSD-----GEQLSA--RDLTMSYALTLTA-----NNRRN 454  
DB 640 AGGSGTGTADGSDLIAGYSGTGTAGSGSSLIAGYSGTGTATSTKSMILTAGYSGTGTAREH 699  
QY 455 SVVPASWGETSASVPGCTCAATSAIGTYSVTVTSMPSIVATG-GTTTATP-----TGS 508  
DB 700 SDLVGTGSGTSTAG-----SNSSLIAGYSGTGTAGFSGSILTAGYSGTGTAGERTSLVAGY 754  
QY 509 GSVTSTG-----KTATATASKTSTTT 528

DB 755 GS-TSTAGYSSSLIAGYSGTGTAGRESTLT 783

RESULT 32

T41144

hypothetical serine rich protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 08-Dec-2000

C:Accession: T41144; T41591

R:Hilbert, H.; Duesterhoeft, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, October 1998

A:Reference number: Z21973

A:Accession: T41144

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1131 <HL>

A:Cross-references: EMBL:AL031907; NID:g3766362; PIDN:CAA21415.1; PID:g3766363; GSPDB:GN

A:Experimental source: strain 972h-; cosmid c18

R:Murphy, L.; Harris, D.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, September 1998

A:Reference number: Z21968

A:Accession: T41591

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 44-1131 <MOR>

A:Cross-references: EMBL:AL031543; NID:g4239667; PIDN:CAA20837.1; PID:g3581899; GSPDB:GN

A:Experimental source: strain 972h-; cosmid c74

C:Comment: For a possible alternative initiator product, see PIR:T42367.

C:Gene: SPOC74.07C; SPDB:SPCC18.01C

A:Map position: 3

C:Keywords: alternative initiators

Query Match 5.5%; Score 153; DB 2; Length 1131;  
Best Local Similarity 20.4%; Pred. No. 0.034;  
Matches 112; Conservative 85; Mismatches 199; Indels 154; Gaps 22;

QY 32 SNEATVARTALINNIGADAWSGADSGIVVASFSTDNPYFYTWTRDGLVLTGLVDF 91  
DB 560 STEALSNLSSTSTASSTSYSPASSYEVANSNSD-----YVS-----QTVSSIT 606  
QY 92 RNQDTSILSTIENYISQAIVOGISNPGDLSGALGPKNVD-----ET 138  
DB 607 ASGTTSTSEI-----VSTPASNSNGSLNGTSSFFVNSGSSQGTPTSSS 654  
QY 139 AYVSGSGRPDRDPALRATAMIGFQWLDNGYSTATDIPWLVNNDLSYVAQYWNQGY 198  
DB 655 STTGSQSLKETSPAL-VSSTVSTSSYDSSSTYNT-----GSSSSDSQSFSGTT 705  
QY 199 YD-----LMEEV-----NGSSFTIAVOHRLVBSAATAVG 232  
DB 706 YSDPTTITSEVSEGLISPTSMQSVSRPQSGDAGFTITFISQSDGFTSGYTSS 765  
QY 233 SCWCDSQAPELLCTYQSFWTGSFILANPDSRSGDANTLGSIHT-EDPEACDDSTF 251  
DB 766 NSQNSGASEPQ-----TAFSSSSSATPTTOSISISTSVSSQSNMSSVS 810  
QY 292 QPCSPALANHKEVDSF-RSIYTLNDGLDSEAVAVGRYPEDITYNCPWFLCTLAALAE 350  
DB 811 SPILSSNSVSTSTIISIASSTTSPIS-----SIAS 844  
QY 351 QLYDA-----LYQWDKQGSLEVTVDLDFPKALYSDAATGTYSSSSTYSIYDAVKTE 405  
DB 845 SPFDAGFTSIYVGTAG-----FSSSF--ALANSSESGADVLSTI-----AKPTF- 890  
QY 406 DGVSVLVERTHASNGSMSEQYDKSDGEQLSARLUTSYALALTANRRNSVVPASNGETS 465  
DB 891 -----KSTSSGSGTYSIPSS-----SRNEGTTYSNSNIYVTG--STLKSLTSSVS 937  
QY 466 -ASVPGTGAATGAIG-----TVSSVTVTSMPSIVATGTTTATPTPGSGSVTSTKTAT 520  
DB 938 TASSYIASHASSNTLTETPKTFSSSTLS-ESISSINTSLIVKPPSSJSSSTTSGLTSS 996



Qy 521 ASKSTTTRS 530  
Db 997 SSTSPSTRS 1006

## RESULT 33

JC2143

ice nucleation active protein - *Erwinia uredovora* (strain KUIN-3)C:Species: *Erwinia uredovora*

C:Date: 30-Sep-1993 #sequence\_revision 20-Aug-1994 #text\_change 17-Mar-1999

C:Accession: JC2143

R:Michigami, Y.; McCabe, S.; Abe, K.; Obata, H.; Arai, S.

Biosci. Biotechnol. Biochem. 58, 762-764, 1994

A:Title: Cloning and sequencing of an ice nucleation active gene of *Erwinia uredovora*.

A:Reference number: JC2143; PMID:94264407; PMID:7764866

A:Accession: JC2143

A:Molecule type: DNA

A:Residues: 1-1034 &lt;MIC&gt;

C:Comment: This protein consists of 52 repeats of closely related 16-amino acid motifs

C:Superfamily: ice nucleation protein

F:161-993/Region: R-domain

Query Match 5.5%; Score 151.5; DB 2; Length 1034;  
Best Local Similarity 21.9%; Pred. No. 0.038;  
Matches 127; Conservative 60; Mismatches 210; Indels 183; Gaps 25;

Qy 47 GADGAVSGADSGIVVASTNDPDTFYWTRSGLVLTUPLFRN-----GDTSL 99  
Db 299 GADSSLIAGVSTQTAGESTQTAGVSTQTAGK-----SDLTAGVSTGAGDS-- 350  
Qy 100 STIENTISAQAIYQGISNPSG-----DLSSG-----AGIGEPKFN 134  
Db 351 SLIAGVSTQTAGEDSLTAGVSTQTAGKSGDLTAGVSTGAGADSLIAGVSTQTA 410  
Qy 135 VDETAVTGWGRPQRDGPALRATMTGFCQWLLDNGYSTATDIWPLVENDLSYVAQY 194  
Db 411 GEESTQTAGVSTQ-----TAQKSD-----LTAGVSTGT-----AGDDSLIA-- 450  
Qy 195 NQGYLMEEVNSSFFITIAVQHRALVEGSAFATVAGSSGSCWCDSPAPILCY--LQSF 252  
Db 451 ---GYSTQTAGDSSLTAGVSTQTAGKSGDLTAGVSTGAGVSSSLIAGVSTQTAG 507  
Qy 253 TGSFIANFDSRSRSGDANTL---GSIHTFDEPAACDDSTFQPCSPRALNHKEVDS- 308  
Db 508 YGSTLTAGVSTQTAGNESDLITGYSTGSTR---AGANSLIAGVSTQTAGSINSLTAG 563  
Qy 309 FRSIYTLNDGLSDSEAVAVGRPEDTYYNGNPFWFLCTLAABQLYDALYQWDKGSL 368  
Db 564 YGSTQTAGREG-SDLT-----GYGSTQTA 586  
Qy 369 DVSLEDFKALYSDAATGYSSSSSTYS--IYDAVKTFADGFVSIYETHAASNGSBOY 426  
Db 587 QEN-----SDLTGYGSTGAGYSSSLIAGVSTQTAGYSIL---TAGVSTQTAG 635  
Qy 427 DKSD-----GEOLAR---DLTWYAL 446  
Db 636 ERSDLTGCGSTAGADSLIAGVSTQTAGVNSILTAGVSTQTAGHNSDLITGYST 695  
Qy 447 LTANNRNSVVPASWGETSASVPGTCA-----TSAIGYSSVTVTSW-PSIV 494  
Db 696 STAGVSSSLI--AGVSTQTAGFKSLTMAGYSSQTAGREOSSLTAGVSTGSMAGYSSLI 753  
Qy 495 ATGCTTTTA-----TPGSGSV-TSTSKTTATASKSTTT 528  
Db 754 AGVSTQTAGVSTLTAGVSTQTAGHNSLTLAGVSTAT 793

## RESULT 34

S25370

MSB2 protein - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: protein G4017; protein YGR014w

C:Species: *Saccharomyces cerevisiae*

C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 21-Jul-2000

C:Accession: S25370; S64305  
R: Bender, A.; Pringle, J.R.

Yeast 8, 315-323, 1992

A:Title: A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence defect.

A:Reference number: S25370; PMID:92383951; PMID:1514328

A:Accession: S25370

A:Molecule type: DNA

A:Residues: 1-1306 &lt;BEN&gt;

A:Cross-references: GB:M7354; NID:g171993; PIDN:AAA4796.1; PID:g171994

R:Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64305

A:Accession: S64305

A:Molecule type: DNA

A:Residues: 1-1306 &lt;RIP&gt;

A:Cross-references: EMBL:Z72799; NID:g1322977; PID:g1322978; MIPS:YGR014w

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:MSB2

A:Cross-references: SGD:S0003246; MIPS:YGR014w

A:Map position: 7R

C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

C:Keywords: transmembrane protein

F:3-19/Domain: transmembrane #status predicted &lt;TM1&gt;

F:1189-1205/Domain: transmembrane #status predicted &lt;TM2&gt;

Query Match 5.5%; Score 151.5; DB 2; Length 1306;  
Best Local Similarity 22.9%; Pred. No. 0.053;  
Matches 134; Conservative 83; Mismatches 240; Indels 129; Gaps 24;

Qy 25 ATLDSWLSNEATYARTAI-LNNIGADG-----VWGSADSGIVVASTNDPDTFYWTRD 79  
Db 267 APLQTESSSFTTASALVSSSTDVDSASPVSSAAGQIASSTSDNFTMSEFTSLT 326  
Qy 80 SGLVLTUPLFRNGDTSLLSTIENTISAQAIYQGISNPSGDLSSGAGLGEPRFNVEDTA 139  
Db 327 S-----TEVD---GSDVS---STVSALLSAPFLQTSNFSFISPVSVFVPSQSSSD-- 373  
Qy 140 YTGSGWRPQRDGPALRATMTGFCQWLLDNGYSTATDIWPLVENDLSYVAQY 199  
Db 374 -----VASSTANVSSPSFDIPQSTGSAV-----SVAQSASALAP 412  
Qy 200 DLMEEVNGSS-----PFTIAVQHRALVEGSAFATVAGSSGSCWCD-----SQAP 245  
Db 413 QSTTEYVYASASTMSGLSTSLQSTTL-DSSSLASSASASDLTDYGVSTASIPILS 471  
Qy 246 CYLQSFMTGSEFTL---ANFDSRSRSGD-ANTLIGSIHTFDEPAACDDSTFQPCSPRAL 300  
Db 472 ASRQASTSSSFVSPSVSFFVPSQSSDVASISAPSVS-----SFSYTSIQAG 522  
Qy 301 NHKEVDSFRSIYTLNDGLSDSEAVAVGRY---PEDTYYNGN---PWFLCTLAABQL 352  
Db 523 SSMWNPSSSTIYSSSTGSEESAASTAGATLSSSSTYAGNLQRPSTSLLESQA 582  
Qy 353 -----YDALYQWDKGSLVTDVSLDFPKALYSDA-----TGTVSSSS 391  
Db 583 TSTSAVLASSSVTSPYTTAGCATASLSLISSTIAETSVQVSYSSCTTALOTSSPSSS 642  
Qy 392 -----STYSSIVDAVTFADGFVSIYETHAASNGSBOY-KSDGEOLASRD 438  
Db 643 TTEGSEFTSGRSTSLVVLQWMPSSISSEFSPQTTQNMASASSSQYTTISSTILQVSD 702  
Qy 439 LTVSYALTLTANNRNSV--VPASWGETSASVPGTCAATSAIGYSS-----VTV 487  
Db 703 TSVSYT---TSSSASVQVSDTPVSY-TTSSSSASVQVSDTPVSYTSSSSASVQVSDTPVSY 758  
Qy 488 TSWPSIVATGCTTTTATPGSGSVTSTSKTTATASKSTTTTSSGMS 533  
Db 759 TTSSSSVQVSDTPVSYTTSSSSVQVSDT---SVPTSSSSASVS 800

## RESULT 35

A46954



## RESULT 37

566852

hypothetical protein YOL155C - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein O0419; protein AOF1001

C:Species: Saccharomyces cerevisiae

C:Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 19-Apr-2002

C:Accession: S66852; S66854; S67325; S70380

R:Arino, J.; Caemmerer, A.; Gamo, F.; Gancedo, C.; Lafuente, M.J.; Aldea, M.; Caas, C.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S66814

A:Accession: S66852

A:Molecule type: DNA

A:Residues: 1-967 &lt;ARI&gt;

A:Cross-references: EMBL:Z74897; NID:g1420063; PID:e251930; PID:g1420064; MIPS:YOL155C

A:Experimental source: strain S288C

R:Gallion, L.; Dujon, B.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S66854

A:Accession: S66854

A:Molecule type: DNA

A:Residues: 1-967 &lt;GAI&gt;

A:Cross-references: EMBL:Z74897; NID:g1420063; PID:e251930; PID:g1420064; MIPS:YOL155C

A:Experimental source: strain S288C

R:Gamo, F.J.; Lafuente, M.J.; Caemmerer, A.; Aldea, M.; Caas, C.; Arto, J.; Herrero, E.

submitted to the EMBL Data Library, July 1995

A:Description: Analysis of the DNA sequence of a 1500 bp fragment of the left arm of ch

pen reading frames.

A:Reference number: S67324

A:Accession: S67325

A:Molecule type: DNA

A:Residues: 1-164; 'STSTSGSSATSGSSSVSGSTATSGSSASAGS', 166-186, 'V', 188-967 &lt;GAM&gt;

A:Cross-references: EMBL:X89715; NID:g1177620; PID:e190152; PID:g1177622

R:Gamo, F.J.; Lafuente, M.J.; Caemmerer, A.; Arino, J.; Aldea, M.; Caas, C.; Herrero, E.

Yeast 12, 709-714, 1996

A:Title: Analysis of the DNA sequence of a 15,500 bp fragment near the left telomere of

A:Reference number: S70379; KUID:96405919; PMID:8810044

A:Accession: S70380

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 55-164; 'STSTSGSSATSGSSSVSGSTATSGSSASAGS', 166-186, 'V', 188-286; 745-773; 781

A:Cross-references: EMBL:X89715

C:Genetics: SGD:S0005515

A:Cross-references: SGD:S0005515

A:Map position: 15L

A:Note: YOL155C

Query Match 5.4%; Score 149; DB 2; Length 967;

Best Local Similarity 20.4%; Pred. No. 0.051;

Matches 117; Conservative 55; Mismatches 201; Indels 200; Gaps 20;

10 SGLVGLAVNISKRAATLDSWLSNEATVARTAILNINIGADGAVWG---ADSGIV--- 61

434 SGLISLSESTGCEVTFSPYSNGAFSPFNALIN-----GGSVSGLQRAEBSGVNNGE 488

62 -----VASPSTDN-----PDYFTWT---RDSGLVLTIVLDFRNG 94

489 INLENGSTVYVVEPVGSGTINIIISGLNLYHVPDTFTGQVVEKGGVL---AVDPTEIN 545

95 DTSLL---STIENYISQAQIVOGIS-----NPSGLSGAGLGEKENVDER 138

546 TPIPVGVYTGEGQIATADVTALSYDSATGVTATQNSQFSFISIGTGFSSSGFVNSEG 605

139 AYTSGWRGPRQDPALRATAMIGFGQWLDNGYTSATDITWPLVRNDLSYVAQVNNQTS 198

606 TFGAIVAYIYNTGCVAVASATPS-----STSTSGATN----- 638

199 YDLMEVNGSSFTTIAVQHRALVGSFAFATAVGSSGWCDSQAPELLICVLOSFWTGSFIL 258

639 -----STSGSTSGFAGV-----TGSTAS 656

QY 259 ANPDSRSRGKANTLLGSIHTEPDAACDSTFQPCSPALANKHEVDSFRSIYTLNDG 318

DB 657 TSFGASVTGSTASTLISG-----SP-----SVTTTIL 684

QY 319 LSDSEAVAGRYPEDTYNGNFWFLCTTAAAEQYDALYQMDKQSLVTDVSLDFKAL 378

DB 685 YATTSTVAVSCSETTDSNGNVYITITTPCSTTATITSCBETGCHVTTST----- 726

QY 379 YSDAATGYSSSSSTYSIYDAVKTRADPFVSIYVTHASNGSMSEQYDKSDGQSLARD 438

DB 737 -GTVAETVSSKS-----YTTVTYVHCNNGCNITVTSCEPEESATP 779

QY 439 LK-WSYALLTA---NNRNSVVPASWGETSASVPGCAATSAIGTSVTVSWPSIV 494

DB 780 TSFKSTVTVYVHCNNGCNT-----KTVSSEAEATTTVSPRTYTTATVTCDD-- 830

QY 495 ANGTGTTATPFGSGSVTSSTKTATVASKTSTT 527

DB 831 -NGCSTKVTSEAPKETSETSETSA-APKTYTT 861

## RESULT 38

S64507

probable membrane protein YGR189C - Yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein G7553

C:Species: Saccharomyces cerevisiae

C:Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 19-Apr-2002

C:Accession: S64507

R:Arroyo, J.; Garcia-Gonzalez, M.; Garcia-Saez, M.I.; Sanchez-Perez, M.; Nombela, C.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64499

A:Accession: S64507

A:Molecule type: DNA

A:Residues: 1-507 &lt;NR&gt;

A:Cross-references: EMBL:Z72974; NID:g1323335; PID:e243566; PID:g1323336; GSPDB:GN00007;

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:CRH1; MIPS:YGR189C

A:Cross-references: SGD:S0003421

A:Map position: 7R

A:Keywords: transmembrane protein

F/6-22/Domain: transmembrane #status predicted &lt;TM&gt;

Query Match 5.3%; Score 147; DB 2; Length 507;

Best Local Similarity 20.8%; Pred. No. 0.028;

Matches 125; Conservative 75; Mismatches 143; Indels 258; Gaps 29;

17 LANVSKRAATLDSWLSNEATVARTAILNINIGADGAVWGADSGIVVASPSTNDPYFTW 76

DB 6 LTVLSASSLSLTFPAAEESTAT-----ADSTTAASTTASCP----- 42

QY 77 TRDSGLVLTIVLDFRNG---DTSLSLTENYISQAQIVOGISNPSGLSGAGLGEPKF 133

DB 43 -----LKT-----TGCTPTAL-----ATSFEDFSSSKWFDTLGHAGEIKY 80

QY 134 NNDETAYTSGWRGPRQDPALRATAMIGFGQWLDNGYTSATDITWPLVRNDLSYVAQY 193

DB 81 GSDGLSMTLA---KRDNPSLKSNFYIMYK-----LEVILKA 115

QY 194 KNOTGYDLMEVNGSSFTTIAVQHRALVGSFAFATAVGSSGWCDSQAPELLICVLOS-- 250

DB 116 ANGTGI-----VSSF-----GKANTLLGSIHTEPDAACDSTF- 291

QY 251 -----FWTG---SFLANPDSRS-----YLOSDDL 132

DB 133 DEIDIEFWAGDNTQFQSNFFSKDITTYDRCGFHGVDTPT--DKFHNYYTLDMAMDXTWY 190

QY 292 -QPCSPRALANKHEVDSFRSIYTLNDGLSDSEAVAVGVYPEDTY-----NGNP-- 340

DB 191 LDGESVRLVSN-----TSSEG-----YPSPMYLMGIMAGGDPDNA 227

QY 341 -----WFLCTLAAAEQY--DALYQMDKQSLVTDVSLDFKALYSDAATG---TYSSES 391

Db 228 AGTEW-----AGETNYNDAPF-----TMYEKIVLDYSTGKITYGDOS 269  
Qy 392 STYSIVDAVKTPADGFVSIVETHAASNGMSBOYDK-----SDPEOLASARDL---T 440  
Db 270 GSWESIE-----ADG-----GSYGNIDQAQEDPAVLANSGSISSSTSSST 311  
Qy 441 WSYAALLTANNRRNSVVPASWGETSASSVPTCATSAIGTYSV-----TTSWPSIVA 495  
Db 312 VSSASSTVSSSVSTVSSASSSTVSSSVSSSSSVSSSSSTSPSSSTATSKTLAS 371  
Qy 496 TGGTT-----TTAPRTGSGSVTSTKATATASKTS-----TTTSGM 532  
Db 372 SSTTSSSISFEKQSSSSSKTVAASSTSESIISSTIKPATVSTTRKVAAPTQOSSV 431  
Qy 533 S 533  
Db 432 S 432

RESULT 39  
C90419  
glucan 1,4 alpha glucosidase (Glucosamylase) [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 07-Mar-2003  
C:Accession: C90419  
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.U.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, H.  
arrett, R.A.; Ragan, M.A.; Seneen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A:Description: Sulfolobus solfataricus complete genome.  
A:Reference number: A99139  
A:Accession: C90419  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-612 <KUR>  
A:Cross-references: GB:AE006641; NID:gl3815775; PIDN:AAK42610.1; GSPDB:GN0155  
C:Gene: SS02473  
C:Superfamily: Sulfolobus type glucan 1,4 alpha glucosidase

Query Match 5.3%; Score 146; DB 2; Length 612;  
Best Local Similarity 21.0%; Pred. No. 0.043; Indels 142; Gaps 23;  
Matches 101; Conservative 66; Mismatches 173;

Qy 16 GLANVISKRATL-----DSWLSN-----EATVARTAIL--NNIGADGAWVSGA 56  
Db 213 GLSDYVKRTPALFKKQDYMWRLMSKNDYGEYDILRRSLILQSHVQNNGAIYASL 272  
Qy 57 DSGIVASPTDNDPIYTTTRDSGLVLTIVDLFRNGDTSLSLTENYISAQAIYVGIS 116  
Db 273 DTIMKFNBDT---YVVVHRDAVFCILAL-----ELMGYFDRSRQFFETRLP 319  
Qy 117 NPSGDDSSGAGLGEPRKENVDETAATGSMGRP-----ORDPALTATAM-IGFGQWL 167  
Db 320 TINGALPH-----KTTVD--GHFGSTWHPWTLDIYPIQEDETALVYLMWPHFSKW-K 369  
Qy 168 DNGYSTATDIWPLVENDLSYVAQYWNQ--TG-----YDLMEEVNGSSFITIAVOHRALV 221  
Db 370 DVDPFKT---YRPMVKGADFLVYRREKATGLPSPFLMEERIGHFHTTTIVINGLR 426  
Qy 222 EGSAPFATVAVSS--CSWCDQAPETILCYLOSFWTGSTILANFDSNRGK-----DAN 271  
Db 427 AADDFAYFGEDELAKQREYVADQMRNSLDLFWVGDH-YARTIYMKGQVHKIDETVSS 485  
Qy 272 TLGSIHTPDPEAACDSTPOCSPRALANHKEVVDSFRSIVTLNDGLDSENAVAGRP 331  
Db 486 ILAAPIENVLP--MSDSRF-----VKDLEVIKLSVKGGLV--RYE 523  
Qy 332 EDTYYY-----NGNPFCLTAAAEQYLDALYQMDKQGLLEVTVDSLDFKALYSDAATGY 387  
Db 524 GDQYLRGNNNSNIWFIISTIMLS-QVYSIMGEKDK----- 556  
Qy 388 SSSSSTYSIVDAVKTPADGFVSIVETHAASNGMSBOYDKSGEOLASARDLTWSYALL 447

Db 557 -----AAEKIDWVLSKSLPTGVIPQIDJDDND-KYPSVPLAMSHBELI 598  
Qy 448 TA 449  
Db 599 RA 600

RESULT 40  
F86719  
hypothetical protein ynhg [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C:Accession: F86719  
R:Boletijn, A.; Wincker, P.; Mager, S.; Jaillon, O.; Malarme, K.; Weisenbach, J.; Ehrlich  
Genome Res. 11, 731-753, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssi  
A:Reference number: A86625; MWID:21235186; PMID:11337471  
A:Accession: F86719  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-614 <STO>  
A:Cross-references: GB:AE005176; PID:gl2723675; PIDN:AAK04856.1; GSPDB:GN0146  
C:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: ynhg

Query Match 5.2%; Score 144.5; DB 2; Length 614;  
Best Local Similarity 21.9%; Pred. No. 0.054;  
Matches 138; Conservative 89; Mismatches 225; Indels 179; Gaps 34;

Qy 5 SLIASGLV-CTGLANVISKRATLDSWLSN--EATVARTAILNNIGADGAWVSGADSGIV 61  
Db 16 SLIALSTLAGFSGLATVNAKTTVADVQVNGEVAIYAQSAAGNQNPQO-----NNGLV 68  
Qy 62 VASPTDNDPIYTTTRDSGLVLTIVDLFRNGDTSLSLTENYISAQAIYVGISNPSGD 121  
Db 69 ---DSDGNP-----VNDSGIANNQ-----TGQAGNTGSKNDGSOQTTLBS--STPS-- 111  
Qy 122 LSSGAGLGEPRKENVDETAATGYSWGRFPORDGPALRATAMIGFGQWLDNGYSTATDIWVP 181  
Db 112 ---TMAVTFPA--TKYVGT-----GVPRGYVDPSTDTTS--VAP 145  
Qy 182 LVENDLSYVAQYWNQ--GYDLWEE--VNGSSE--FTIAVOHRALVEG-----SA 225  
Db 146 -VTTDASGLADFLGTLIDGYLLFHQVTTVNGITVGFIVOVSHEDSQAGIVNYPRLDMS 204  
Qy 226 FATAVSSCSW-----CDQAPETILCYLOSFWTGSTILANFDS-----RSGKD 269  
Db 205 SSAGLGTSATTNADNPNQGTPEQOIANPNATGNSDQTLTNVTDNAGNENLANGTWTNGSD 264  
Qy 270 -----ANTLISLHT-FDPEAACDS-----TFQPCSPRALANHKEVVDS 308  
Db 265 NQMTTAAAGNTVNNKVVNTVFPDSSQTNNGNGTGVGTIVLDQ--LPNNLVNVSSTV--- 319  
Qy 309 FRSIYTLNDGLDSENAVAGRPEDYTY---NGNPFCLTAAAEQYLDALYQMDKQGS 364  
Db 320 --TVSTVIYVTVNGSGTKVGTLLPTTDYITTNQNKIVTLLTTAQQOHAAASLGSADGA 377  
Qy 365 LEV---TDSVLDFFKALYSDAATGYSSSSSTYSIVDAVK--TFADGFVSIVETHAASNG 420  
Db 378 LNIILIPSTYKSAIGSA--TDSATTTITMAGADLSTTAAKSTLVNGLGEMTTDASTNA 435  
Qy 421 SMS-----EQYDKSGEOL-----SARDLTWSYALLTANNRRNSVVP 458  
Db 436 ALAGATFTVVRADNKEDADQFVEANAAFPNNSASGCTVNNLTJSKAAFTYGDTSGNA--- 492  
Qy 459 ASWGETSASSVPTCATSAIGTYSVTVTSWPSIVATGCTTT-----ATPTG--- 507  
Db 493 ---NTSATAPVFTTGKQGIATFNGLNLDV---NNTDGSNTTNYVLVEAAPTYQLP 544  
Qy 508 -----SGSVTSTSKTATATASKTSTTRS 530

Db 545 SVTTAANLTGAVTA---STAPATDTTITNN 572

## RESULT 41

T11678

hypothetical protein SPBC21D10.06c - fission yeast (*Schizosaccharomyces pombe*)  
C:Species: *Schizosaccharomyces pombe*  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999

C:Accession: T11678  
R:Seeger, K.; Harrie, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, September 1998

A:Reference number: Z17313

A:Accession: T11678

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-948 &lt;SEE&gt;

A:Cross-references: EMBL:AL031536; NID:e1319499; PID:e1319505

A:Experimental source: strain 972h(-)

C:Genetics:

A:Map position: IIR

A&gt;Note: SPBC21D10.06c

Query Match 5.2%; Score 144; DB 2; Length 948;  
Best Local Similarity 20.4%; Pred. No. 0.11;  
Matches 114; Conservative 84; Mismatches 234; Indels 126; Gaps 22;

QY 25 ATLDSWLSNEATVARTAILNNGADGAWVSGADSGIVASPTDNDPDTYTRDGLV 84  
Db 211 SLSLSEFVITNVDSSTTSVINIYIGA---STLESSLTNTVSTPSTFETKSSTSVPT 265  
QY 85 KTLVDLFRNGDTSLSLTIENYISAQAVOGISNP---SGDLSSGAGLGEPPNDETAAT 141  
Db 266 QTI-----DSSGFTS-----STPVLSTSSSTSSSSQDSTIIDTPET 304  
QY 142 --GSGCRPORDG-----ALPATMIGFGQMLDNGYSTATDIWPLVRNDLSY 189  
Db 305 IATSTLOPTTSPITTSAPLSGALPTTPSSLSTEVEVEYFTKITL-----DTSS 355  
QY 190 VAQYMNQGYDL---WEVNGSFFITIAVQHRALVGSAR----- 226  
Db 356 IVTY--STVEVETLETETTSSEISIIYNFSTPIGSSFPDGFKPINFTSPSLTSSTK 413  
QY 227 ---ATAVSSGSCWCDQAPFILLCYLQ-SFWTGSFIIANFDSRSGKADANTLGSHTPDP 282  
Db 414 KIPSTLLPRTSSKMITTTTSVSNNQSSFLIISTFTSSYHSPRPVSSVPLTS----- 467  
QY 283 EAACDDSTPOCPSPRALANHKEVVDSPRSIYTLNDGLDSEA---VAAGRYPEDTY-YN 317  
Db 468 -----NPFSSISHSASSILPIPPSSYLSNVTLLHSSVQSSQSFVSV---PSTQSY 518  
QY 338 GNPWFICTLAAAEQLYDALYQMDKQGLEVTQVSLDFPKALYSDAATGYSSSSSTYSST 397  
Db 519 TSSNFTPTTITISL-----SSFTTIVSSSFQYSSLSNVTYTNAAOSSLSSSN 568  
QY 398 VDAVKTFADGFVSIVETHAASNGMSEQYDKSDGEOLASADLTWSY-AALLTANRRNSY 456  
Db 569 SGAHTHSSSIYSSGSSSALSSTIYSSINSSSVIYSSSSLOQSSSVITTTTSGSV 628  
QY 457 VPASWGETSASVVP-GTCAATSAIGTYSVTVTSWPSIYATGG---TTTATPTGS--G 509  
Db 629 -----GFTTITATPVGSTAGTVV---DIPPSWTEETVTSQSVGFTTITATPVSTAG 679  
QY 510 SVTSTSKTATASKSTT 527  
Db 680 TVLVNDIPSPSWTEIVTS 697

## RESULT 42

T30531

agglutinin-like adhesin - yeast (*Candida albicans*)

C:Species: *Candida albicans*

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 17-Mar-2000

C:Accession: T30531

R:Gaur, N.K.; Klotz, S.A.  
Infect. Immun. 65, 5289-5294, 1997

A:Title: Expression, cloning, and characterization of a *Candida albicans* gene, ALA1, that  
A:Reference number: Z20847; PMID:98053977; PMID:9939828

A:Accession: T30531

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1419 &lt;GAP&gt;

A:Cross-references: EMBL:AF025429; NID:g2522218; PID:g2522219; PIDW:AA08083.1

C:Genetics:

A:Gene: ALA1

C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase }

Query Match 5.1%; Score 142; DB 2; Length 1419;  
Best Local Similarity 19.9%; Pred. No. 0.26;  
Matches 134; Conservative 89; Mismatches 210; Indels 242; Gaps 34;

QY 26 TLDSWLSNEATVARTAILNNGADGAWVSGADSGIVASPTDNDPDTYTRDGLV 83  
Db 440 TQGFV--SESFTSTTITNSL-----KGTDS-VVREP--HNPVTTEFWMS-ESFAT 486  
QY 84 LKTIVD-----LFRNGDTSLSLTIENYISAQAVOGISN-PSGDLSSGAGLGP--- 131  
Db 487 TETITSKPEGDSVIVREPHNPVTYTEFWMSYATTEITNGPEG--TDSVIVREPHNP 544  
QY 132 -----KFVDETAATYGS-WGRPORD-----GPALPATMIGFGQMLDNGYSTAT 176  
Db 545 YTTTTFKMSAYATTEITINKPEGDSVIVKEPYNPTVTTE---N--SSSYATTE 597  
QY 177 DIWVP-----LVR--NDLSYVAQYMNQGYDLMEVNGSFFITIAVQHRALVGSARA 227  
Db 598 ITNGPEGDSVIVREPHNPVTYTEFWMS--YATTEITNGPEGDSVIVHDLSESSST 656  
QY 228 TAVGSSCWCWCDQAPFILLCYLQSFWTGFTLANFDSRSGKADANTLGSHTDPEACD 287  
Db 657 TAIBESSDNISSSAOE-----SSSVEQSSSIVGLSSSDIPLSSD 697  
QY 288 DSTPOCPSPRALANHKEVVDSPRSIYTLNDGLSDSAVAVAGRYPEDTYNNGNPFCTLA 347  
Db 698 ---NPSS-----STGLTSSSTSVSYDSD--SSSIESSTLS 731  
QY 348 AAEQ---LYDALYQMDKQGS-LEVTDV---SLD----- 373  
Db 732 SDDRCSSTSDTTSFMDSSSDLESTISLWSSSIDQSHLVQSVNSISTQELSSSS 791  
QY 374 -----FPKALYSDAATGYSSSSSTY-----SSIYDAVK 402  
Db 792 ESSSTPATDALVSSDASISLSDTSSYSPSTISSSDDEPHITAGESDLSIFITSV 851  
QY 403 TRADGFVSIVETHAAS-NGMSEQYDKSDGEOLASADLTWSYAAALLTANRRNSVVP--- 458  
Db 852 ISSDS-VSLTSPASSFSDSSSLNSDSSSSPSSDQSD-----ILTSSFTLLVPSPS 903  
QY 459 -----ASWGETSASVPGTCA-----ATSAIGTY 482  
Db 904 LSSSSLSLTYHYNNSTTYHASESSSSVAPSMASEANDTYLLSSTDTTSSIGTD 963  
QY 483 SSYT-----VTSWPS-----IVATGTTTATPTGS-GSVTSTSKTTA 519  
Db 964 SSTVTFCRDNDGDCIVTGMPSSSIDSEQTSVTVTTSSFVASSTPSASQISITDNPIDS 1023  
QY 520 TASKSTTRSGMSL 534  
Db 1024 SOTSASSSTKSSVS 1038

## RESULT 43

I40712

endo-1,4-beta-xylanase (EC 3.2.1.8) D precursor - *Cellulomonas fimi*

N/Alternate names: xylanase D

C:Species: *Cellulomonas fimi*

C:Date: 16-Aug-1996 #sequence\_revision 22-Nov-1996 #text\_change 18-Jun-1999

C:Accession: I40712

R,Millward-Sadler, S.J.; Poole, D.M.; Henriessat, B.; Hazlewood, G.P.; Clarke, J.H.; Glib  
Mol. Microbiol. 11, 375-382, 1994  
A>Title: Evidence for a general role for high-affinity non-catalytic cellulose binding  
A:Reference number: 140712; PMID:9422415; PMID:8170399  
A:Accession: 140712  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-644 <RES>  
A:Cross-references: EMBL:X76729; NID:g558176; PIDN:CA54145.1; PID:g558177  
C:Genetics:  
A:Gene: xymD  
C:Function:  
A>Description: catalyzes the hydrolysis of 1,4-beta-xylosidic bonds in xylans  
A:Pathway: xylan degradation  
C:Superfamily: Cellulomonas endo-1,4-beta-xylanase D; endo-1,4-beta-xylanase homology;  
C:Keywords: extracellular protein; glycosidase; hydrolyase; polysaccharide degradation  
F:1-35/Domain: signal sequence #status predicted <SIG>  
F:36-644/Product: endo-1,4-beta-xylanase D #status predicted <MAT>  
F:54-229/Domain: endo-1,4-beta-xylanase homology <XFL>  
F:361-508/Domain: nodB homology <NODB>  
F:126.216/Active site: Glu #status predicted

Query Match 5.1%; Score 141.5; DB 1; Length 644;  
Best Local Similarity 20.7%; Pred. No. 0.093;  
Matches 142; Conservative 76; Mismatches 238; Indels 229; Gaps 35;

QY 28 DSWLSNEATVARTAI--LANNIGADGAVSGADSGIVASPS---TDN-----PDYFYT- 75  
DB 3 DSEPARTRTRRRRLQALVTGLAAGALVACA---LAASAPAAAVTNTGTGHDGYYSF 59  
QY 76 WTRDSGLVLTVDLFNNGDTSLSTIENTYISAQAIYQGISNPSGDISGAGICEPKFNV 135  
DB 60 WTPSGSV---SMDLNGGGGYTRWSNTGNFVAG---KGMSTGGRKTVSYSG---QFNP 108  
QY 136 DEFAYTSGWRPQRDGPALATAMIGCOMLNDGYSTATD-----IWPVLV 183  
DB 109 SRNAYLTLYGWTQ--SPLVEYIVDSWGYRPTGTFTMGIVTSDGGTYDIYRTORVNPST 166  
QY 184 RNDLSYVAQYWN---QTG-----YDLME-----EYNGSSF 211  
DB 167 EGGSSTFYQYWSVRQORGTGTTSGNHPDAMASKGNLGRHNYMIMATEGYSSSSSI 226  
QY 212 TIAVQHRALVEGSAFA---TAVGSSCWCDSPAPELLCYQSF-----WTGSF 256  
DB 227 TVS-----EGSGCGGGDTGGCGGSGTCSVTATRAEWSDRPNVTYSVSGSAMTVNL 279  
QY 257 ILANFDSRSGKDN-----TLGSIHF-----DEAACDSTPQCS 295  
DB 280 ALNGSQTIIQASWNNVNTGSGSTRVTVPNGSGNTGVTVMKNGSTTPAATCAGSGGCTAT 339  
QY 296 PRALANKEVVDSPRSY---TLNDG-----LSDSEAVAVGRYPEDTYVYNGNP 340  
DB 340 PTPPTPTTPPGSSAGYVGLTTPDGPNTGTNIIISTLYQYGTAL--VPTTQGNAGNP 398  
QY 341 WFLCTLAABQLYD---ALYQWD-----KQGSLEVTDVSLDFP 375  
DB 399 SLM-----QAYKNAGVQIGNSHMDPHLVNMSQSDMOSQLTRTOAIQOTACVPTLIF 451  
QY 376 KALY---SDAATGYSSS-----SSTYSYDAKTPADGVSLVE--- 413  
DB 452 RPPGESNATLROVESSLGREIIVDVSODPMNNSASQIRQASRLTNQIILMHMPA 511  
QY 414 -THAASGMSSEQYDKSDGEOLASRDLTWSYALLTANNRNSVVPAS--WGTSASSVPG 471  
DB 512 ATYQALPGLI-----QDLRSRLCTGHISSTG---RAYAPSSAGGGGGGGGTG 558  
QY 472 TCAATSAIG-----TY-----SSVTY-----SWPSIVATGTTTATPT 506  
DB 559 SCVSVAVRGEEMADRFNVTYSVSGSSSWVYTLGLNGGVSQSWNALTLTSSGTVTARPN 618  
QY 507 GSGS---VT--STKTTATASTKTTT 527  
DB 619 GSGNSFGVTYKNGSSATPGATCAT 643

RESULT 44  
B86807  
hypoetical protein yoiC [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
C/Species: lactococcus lactis subsp. lactis  
C/Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #ext\_change 03-Aug-2001  
C/Accession: B86807  
R/Bolotin, A.; Winkler, P.; Mager, S.; Jalllon, O.; Malarme, K.; Weisenbach, J.; Ehrlich  
Genome Res. 11, 731-753, 2001  
A>Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssi  
A:Reference number: A86625; PMID:21235186; PMID:11337471  
A:Accession: B86807  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1441 <STO>  
A:Cross-references: GB:A8005176; PID:g12724450; PIDN:AAK0556.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: yoiC

Query Match 5.1%; Score 141.5; DB 2; Length 1441;  
Best Local Similarity 19.3%; Pred. No. 0.29;  
Matches 119; Conservative 87; Mismatches 237; Indels 175; Gaps 24;

QY 13 VCTGLANVSKRATLDSWLSNEATVARTAILNNG-----ADGA--WVSG 55  
DB 438 ITQGISNVIOE-----IANNIGALNGNIIVSILITTIANGVLQTVTA 479  
QY 56 ADSGIYVASTNDPNDFYTWTRDS--GLVLTVDLFNNGDTSLSTIENTYISAQ---- 109  
DB 480 VTMGLI-----PTDIVQNFITTVNSAFGSTNISLNGLSGTSVITNALMSITGAQLPII 535  
QY 110 -----AIYQGISNPSGDISGAGICEPKFNVDETAYYTSGWRPQRDGPALRA 156  
DB 536 NQSIGLVGNVSGITQGIOTGQVYS-----LINTLATLNGVAAVQASQDQMRA 588  
QY 157 TA-----MIG---PQWMLI-----DNGYSTAIDYWPV--VRNDLSYVAQYWNQT- 197  
DB 589 LAKEPKVIANAGTFNESFLANTONADGYTVSDTGSHPALTATNKVALVYNGYIQAT 648  
QY 198 -----GYLMEEVNSSFFITAVQHRALVEGSAFAPTAVAGSSCWCDSPAPELLCY 247  
DB 649 SVGLKTIIVTIPQMLIDAIAGLSNPNLGLNTGSLFDG--VENTVINGVNVGVSFASAMQG 707  
QY 248 LOSFWTGSFLIANFDSRSGKDNATLGSIHTEFDEPAACDSTFQCSPRALANKEVVD 307  
DB 708 LNSTVAGASLVNNTSNMTVQTSNIVSQY-----VNSIAN-----A 744  
QY 308 SFRSIYTLNDGLS---DSEAVAVGRYPEDTYVYNGNPWFLCTLAABQLYDALYQMDKQGS 364  
DB 745 SATGVAIVTVGTPAADPSSQIATNLATANHYYQGO-----SAATTOROLDNVPE---T 796  
QY 365 LEVTVDSLDPFKALYSDAATGYSSSSSTYSYVDA--VKTFPDGFYSIETHAASGSM 422  
DB 797 FKVTPTPAL-----AGQINTTDTYSVVOYNTNTIKGSIYVVDSTKALANAK 844  
QY 423 SEQYDKSDGEOLASRDLTWSYALLTANNRNSVVPV----- 459  
DB 845 ATAYEKSDETVAPAPAPAY--VLNDVNNKAVSADANGVINFYVDAIOPTLTKOSTID 903  
QY 460 ---SWGTSASSVPGTCAATSAIGYSSVTVTWSPSIVATGTTTAT---PTGSGSVTS 513  
DB 904 EGSITW--TADNFTGSGTGTIVTINDVTGTVDSTPGVSSTVYTYDPTGASISS 961  
QY 514 TSKTTATASTKTTTSSG 531  
DB 962 VANITVDSSTNTNPTDG 979

RESULT 45  
S11672  
Ice nucleation protein - Xanthomonas campestris

C:Species: Xanthomonas campestris  
C:Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 26-Aug-1999  
C:Accession: S11672  
R:Zhao, J.; Oreeer, C.S.  
Mol. Gen. Genet. 223, 163-166, 1990  
A:Title: Conserved repetition in the ice nucleation gene inax from Xanthomonas campestris  
A:Reference number: S11672; WUID:91080859; PMID:2259339  
A:Accession: S11672  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1567 <ZHA>  
A:Cross-References: EMBL:X52970; NID:g48531; PIDN:CAA37140.1; PID:g48532  
C:Superfamily: ice nucleation protein

Query Match 5.1%; Score 140.5; DB 2; Length 1567;  
Best Local Similarity 23.1%; Pred. No. 0.37;  
Matches 126; Conservative 68; Mismatches 229; Indels 123; Gaps 25;

QY 47 GADGAMV-----SGADSGIVVAPSTD-----NPDYFYWT--RDSGLVLTIV 88  
DB 725 GADSTLIAGVSTQTSGSDSLTAGVSTQTRAKGSDVTAGVSTGTAGADSTLIAGVGS 784  
QY 89 DLFRNGDTLSTIENYISAQAIWGISNPSGDLSSG-----AGLGEPEKENVDETA 139  
DB 785 TQTSGSDSLTA---GYGSTQTRAKGSDITAGVSTGTAGADSTLIAGVSTQTSGSDS 841  
QY 140 YMGSMGRPQ--RDGPALRATAMIGFGW-----LLDNGYSTATDIWPLVRNDLSYV 190  
DB 842 LTAGVSTQTRAKGSDVTA---GYGSTGTAGADSTLIAGVSTQTRAKGSDVTA---AGSDSLT 891  
QY 191 AQYWN-QTGYDLMEEVNCSFPTIAVQHRALVEGSAFATAVGSSCSCMSQAPELICYLQ 249  
DB 892 AGVSTQTRAKGSDVTAGVSTGTAGADSTLIAGVSTQTSGS-----DSSLTAGVSTQ 946  
QY 250 SFWTGSFIANFDSRSGKDANTLLGSIHTFDEAACDDSTFQPCSPRALANHKEVVDSE 309  
DB 947 TARKGSDMTAGVSTGTAGADSTLIAGVSTQTSGSDSLTAGVSTQTRAKGSDVTAGY 1006  
QY 310 RSIYTLNDLSDSEANAVGRYPEDTYNNGNPWFCLTLAAAEQLYDALY---QWDKQGSLE 366  
DB 1007 GS-----TGTAGADSTLIAGVGS-----TGTAGSDSLTAGVSTQTRAKGS-- 1048  
QY 367 VTDVSLDPFKALYSDATGTYSSSS-----STYSIVDAVKTFFADGFVSIIVETHASN 419  
DB 1049 --DVTAGY-----GSTGTAGADSTLIAGVSTGTAGSDSLTAGY-----STQTRAK 1094  
QY 420 GS-MSEQYDKSD-----GEQLSA--RDLTWSYALLLTANNRRNSVVPASWGE 463  
DB 1095 GSDITAGVSTGTAGADSLTAGVSTQTRAGYDSNLTAGYSTQTA--REDSLTAGYGS 1152  
QY 464 TSASSVPGCATSAICTYSSVVTWSPISVATG--CTTTATPTGSGSVTSKTTATAS 522  
DB 1153 TSTAG---HDSLTAGVSTGTAGYNSLTTGGYSTQTADE--SSSLTAGYGSTTAG 1205  
QY 523 KTSSTT 528  
DB 1206 YDSTLT 1211

Search completed: June 28, 2004, 07:45:23  
Job time : 27 secs



**THIS PAGE LEFT BLANK**